

DISEÑO, DESARROLLO Y EVALUACIÓN
DE SITUACIONES PROBLEMÁTICAS
DE ESTADÍSTICA

FELIPE FERNÁNDEZ
OLGA L. MONROY
LILIANA RODRÍGUEZ



una empresa docente®

Bogotá, 1998

PRIMERA EDICIÓN, FEBRERO DE 1998

DISEÑO, DESARROLLO Y EVALUACIÓN DE SITUACIONES PROBLEMÁTICAS DE
ESTADÍSTICA

Felipe Fernández, Olga Lucía Monroy y Liliana Rodríguez

D. R. © 1998 una empresa docente®.

Ninguna parte de esta publicación puede ser reproducida, archivada o transmitida en forma alguna o mediante algún sistema, ya sea electrónico, mecánico, de fotorreproducción, de almacenamiento en memoria o cualquier otro, sin el previo y expreso permiso por escrito de “una empresa docente” y de los autores.

Diseño carátula: una empresa docente®

una empresa docente®

Universidad de los Andes

Cra. 1 Este # 18 A - 70

Apartado Aéreo 4976 Tel.: (57-1) 284-9911 ext. 2717. Fax: 3520466 ext. 2709

Servidor WWW: <http://ued.uniandes.edu.co>

Bogotá. Colombia

Esta publicación fue realizada con el apoyo financiero de Colciencias (BID) y la “Fundación para la promoción de la investigación y la tecnología” del Banco de la República.

ISBN 958-9216-14-7

Impreso en Colombia

TABLA DE CONTENIDO

Presentación	v
--------------------	---

1. INTRODUCCIÓN

Contexto	1
Propuesta de trabajo	4
Aproximación conceptual de la propuesta	5

2. ELEMENTOS PARA EL ANÁLISIS DEL CONTENIDO

Introducción	8
Conocimiento conceptual y procedimental	9
Sistemas de representación	11
Resolución de problemas en tests de hipótesis	12

3. ANÁLISIS DEL CONTENIDO EN LOS TESTS

DE HIPÓTESIS

Introducción	14
Unidades de análisis	14
Conceptos y procedimientos	20
Análisis en términos de representaciones externas	25

4. ELEMENTOS PARA EL ANÁLISIS

DEL APRENDIZAJE

Introducción	26
Comprensión	26
Concepto y concepción	27
Errores y dificultades	29
Sistemas de representación	32

5. ANÁLISIS DEL APRENDIZAJE DE LOS TESTS

DE HIPÓTESIS

Introducción	34
Comprensión estocástica	34
Análisis en términos de representaciones	49

6. ELEMENTOS PARA EL ANÁLISIS DE LA ENSEÑANZA

Introducción	51
--------------------	----

Análisis de textos	51
Situaciones problemáticas y enseñanza	52

7. ANÁLISIS DE ALGUNOS ASPECTOS DE LA ENSEÑANZA EN LOS TESTS DE HIPÓTESIS

Introducción	55
Análisis de los textos	55
Problemas de estadística y su enseñanza	62

8. SITUACIÓN PROBLEMÁTICA N° 1: INFERENCIA ACERCA DE LA PROPORCIÓN DE UNA POBLACIÓN

Introducción	65
Análisis del contenido estadístico	65
Análisis de aprendizaje	70
Análisis de enseñanza	86
Evaluación	88

9. SITUACIÓN PROBLEMÁTICA N° 2: INFERENCIA ACERCA DE LA DIFERENCIA DE MEDIAS DE DOS POBLACIONES

Introducción	94
Análisis del contenido	94
Análisis del aprendizaje	100
Análisis de enseñanza	114
Evaluación	118

10. SITUACIÓN PROBLEMÁTICA N° 3: INFERENCIA ACERCA DEL COEFICIENTE DE CORRELACIÓN DE UNA POBLACIÓN

Introducción	127
Análisis de contenido	127
Análisis del aprendizaje	132
Análisis de enseñanza	150
Evaluación	152

11. SÍNTESIS Y CONCLUSIONES 162

12. REFERENCIAS 168

PRESENTACIÓN

Este libro es resultado de un proyecto de investigación en educación estadística. En él se presenta el proceso que se siguió para la realización del diseño, la aplicación, la observación y el análisis de los resultados de tres situaciones problemáticas de estadística de tests de hipótesis. Dichas situaciones se propusieron en el contexto de un curso de estadística para ciencias sociales durante el primer semestre de 1997.

El libro está dividido en doce capítulos. En el primero, se presenta una introducción en la que se exponen algunas consideraciones acerca del contexto de la investigación. En el segundo, se presentan los elementos conceptuales con base en los cuales se plantean de manera general el análisis del contenido y luego, en el tercer capítulo, se presenta el análisis realizado con base en los elementos conceptuales presentados en el segundo capítulo; en el cuarto y quinto capítulo, se presentan los elementos conceptuales que se tuvieron en cuenta para el análisis del aprendizaje de los estudiantes y el análisis realizado con base en los conceptos presentados; similarmente, en los capítulos sexto y séptimo se presentan respectivamente los elementos conceptuales que se tuvieron en cuenta para el análisis de la enseñanza y el análisis realizado con base en estos elementos. El libro termina con un capítulo de síntesis y conclusiones y con la lista de referencias de la literatura en la que se apoyó el trabajo realizado.

Los análisis que se refieren a cada situación problemática particular, se presentan en los capítulos octavo, noveno y décimo. Además, cada una de las tres situaciones problemáticas que se consideraron contienen secciones en donde se hace referencia a aspectos tales como:

- Posibles tipos de situaciones en las que se pueden poner en juego los conceptos de interés propios de cada elección temática.
- Ejemplos de situaciones en donde se explicitan los conceptos y procedimientos que se ponen en juego.
- Descripciones de los tipos de errores que se podrían identificar en los estudiante.
- Comentarios acerca de las posibles dificultades que se pueden encontrar detrás de los errores, donde incluso en algunos casos, se alcanza a insinuar la existencia de algunos obstáculos de conocimientos que los estudiantes tienen cuando están resolviendo problemas de este tipo.

Pensamos que el lector interesado en el tema de la educación estadística puede encontrar en este libro un valioso instrumento de apoyo para su labor docente o prarainiciar procesos de investigación en este campo.

1. INTRODUCCIÓN

CONTEXTO

Contexto académico

Esta investigación se realizó en el contexto de un curso de estadística de nivel universitario llamado Estadística 2. Dicho curso es ofrecido por el Departamento de Matemáticas de la Universidad de los Andes a estudiantes de ciencia política, antropología y lenguas modernas. El trabajo se realizó durante el segundo semestre de 1996 y el primero de 1997 por tres profesores-investigadores que estuvimos encargados de la realización de este proyecto. La tarea principal consistió en diseñar, aplicar y evaluar un conjunto de situaciones problemáticas de estadística para el curso de estadística mencionado antes.

Estadística 2 es la continuación de un primer curso de estocástica¹ llamado Estadística 1. En este primer curso de estadística los estudiantes tienen, en general, su primer contacto con los temas de probabilidad y estadística. La Tabla N° 1 que se presenta a continuación contiene los temas que se han trabajado en estos dos cursos durante los semestres de 1995 y 1996.

*Tabla N° 1. Contenidos de los cursos
de Estadística 1 y Estadística 2*

Temas	Curso
Estadística descriptiva Distribución de frecuencias y representaciones gráficas Medidas de tendencia central (media, mediana, moda) Medidas de variabilidad (rango, varianza, desviación típica) Medidas de posición (percentiles, z-scores)	Estadística 1
Teoría de la probabilidad Leyes (adición, multiplicación) Sucesos independientes e incompatibles Variables aleatorias Distribuciones de probabilidad Distribuciones binomial y normal Teorema del límite central	Estadística 1

1. Término utilizado para referirse a la estadística y la probabilidad simultáneamente.

*Tabla N° 1. Contenidos de los cursos
de Estadística 1 y Estadística 2*

Temas	Curso
Inferencia estadística Estimación de parámetros (media, diferencia de medias, varianza, proporción, coeficiente de correlación) Tests de hipótesis (medias, varianza, proporción, correlación, independencia, bondad de ajuste, análisis de varianza)	Estadística 2

En la tabla se puede ver que mientras que en el curso de Estadística 1 se desarrollan casi todos los temas que se refieren a las partes de estadística descriptiva y teoría de la probabilidad, en el de Estadística 2 se hace más énfasis en la estimación y en particular en los tests de hipótesis. Esta tabla prácticamente coincide con la que Garfield (1995), presenta cuando se refiere a los temas de estadística y probabilidad que se tratan en cursos introductorios de nivel universitario.

Antecedentes

Los cursos de Estadística 1 y Estadística 2 han sido objeto de múltiples reformas a lo largo de los últimos años. Sus propuestas curriculares han evolucionado y en este sentido queremos hacer una pequeña reseña de cómo ha sido este proceso.

A comienzos de 1987, un grupo de profesores de la Universidad, desearon introducir cambios que mejoraran la situación de estos cursos concibieron un proyecto de innovación curricular llamado “Matemáticas y Sociedad” (Gómez, 1993). Como parte de los resultados de este proyecto fueron producidos tres textos (Perry et al., 1990; Fernández et al., 1995 y Perry et al., 1996) relacionados con temas de la estocástica. En estos textos y también en algunos otros materiales, como talleres de trabajo para la clase y trabajos de proyectos de investigación de los estudiantes, se puede observar la intención del proyecto de innovación curricular, de establecer metas que apunten a que los estudiantes aprecien el papel vital que juega la estadística en la investigación científica y a que aprendan a utilizar los métodos estadísticos más conocidos y aplicados en la práctica. En este sentido también compartíamos las metas comunes que según Gal y Garfield (1997) deben ser comunes a todos los niveles y contextos educativos. Estas metas establecen que

[...] al terminar los estudiantes sus encuentros con la estadística, lleguen a ser ciudadanos informados capaces de:

- Comprender y tratar con incertidumbre, variabilidad, e información estadística en el mundo alrededor de ellos, y participen efectivamente en una sociedad llena de información.

- Contribuir a, o tomar parte en la producción, interpretación y comunicación de datos pertenecientes a problemas que se encuentren en su vida profesional.

El establecimiento más consciente de metas era pues, una de las características del proyecto que en ese entonces estábamos iniciando. Es claro, como Hogg (1990) lo menciona, que todos los cursos introductorios de estadística deben establecer cuidadosamente sus metas, pues son éstas las que permiten a los profesores evaluar más efectivamente las propuestas curriculares que quieran desarrollar. Sin embargo, muchos cursos utilizan lo que Macnaughton (1996) ha llamado metas “basadas en tópicos”. Cuando se piensa así, las metas simplemente especifican una serie de temas que deben ser vistos. Se podría decir que así era la situación de los cursos de Estadística 1 y 2, a mediados de los años 80.

Desde cuando fue elaborado el proyecto de innovación curricular mencionado más atrás, la coordinación de Estadística 1 y Estadística 2 estuvo a cargo de “una empresa docente”² hasta el segundo semestre de 1995. Luego, desde 1996, la coordinación volvió a estar a cargo del Departamento de Matemáticas y como consecuencia de este cambio nuevamente se produjeron sustituciones en los textos y algunas variaciones en las metodologías de enseñanza. Sin embargo, el interés por proponer nuevos problemas y situaciones didácticas a los estudiantes siguió vigente y con ello también, el apoyo a este proyecto.

A pesar de los cambios sucedidos en el último cambio de coordinación, siguieron existiendo dentro de las actividades de los cursos de Estadística 1 y Estadística 2, además de los parciales y de los exámenes finales, espacios de trabajo donde los estudiantes se debían enfrentar con problemas de estadística, que pusieran en juego los conocimientos aprendidos. A este último tipo de actividades usualmente se les ha llamado talleres.

Fue así como, una revisión crítica a los parciales, exámenes y talleres que se les venía proponiendo a los estudiantes, dejó entrever que en general, no eran coherentes con la posición de los autores acerca de cómo se debía enseñar estadística (Fernández, 1996). Entonces ésta fue, entre otras, una de las razones por las que en este proyecto se presentó una propuesta metodológica para llevar a cabo el diseño, la aplicación y la evaluación de talleres, que de ahora en adelante vamos a referir con el nombre de situaciones problemáticas de estadística.

2. Al poco tiempo de iniciar el proyecto de innovación curricular, los profesores que formularon la propuesta constituyeron al interior de la Facultad de Ciencias un grupo de trabajo llamado “una empresa docente”, que en la actualidad es un centro de investigación en educación matemática.

PROPUESTA DE TRABAJO

Propósitos

Si tenemos en cuenta no sólo la crítica que se ha hecho al contenido mismo de las evaluaciones escritas (parciales, exámenes y situaciones problemáticas) que se les propone a los estudiantes, sino también los análisis de las respuestas que ellos dan durante el desarrollo del curso y al finalizar el semestre, encontramos por un lado, que los estudiantes tienen múltiples dificultades y concepciones del conocimiento estadístico que normativamente se consideran incorrectas y que persisten al finalizar el semestre de estudio. Por otro lado, también encontramos que hace falta un trabajo más intenso que apunte a la elaboración de situaciones problemáticas que sean propicias para poner en evidencia de manera más consciente las concepciones erróneas de los estudiantes y que a la vez dispongan de metodologías de evaluación para que éstas puedan ser superadas.

Este libro tiene el propósito de mostrar la manera como se llevó a cabo dicha propuesta, la cual persigue que en dichas situaciones problemáticas:

- Se pongan en evidencia concepciones erróneas.
- Se pueda indagar acerca de las posibles causas de las concepciones erróneas.
- Se genere un ambiente más motivante de aprendizaje y enseñanza entre los estudiantes y entre los estudiantes y el profesor para superarlas.
- Se aproveche la experiencia de estos diseños para que los profesores investigadores, la apliquen en su futura práctica docente.
- Se propongan y se indague acerca de preguntas de investigación alrededor del tema, con base en los diseños propuestos.

Concretamente, en las situaciones problemáticas de estadística que se diseñaron se consideraron los siguientes temas: diferencia de medias, correlación cuantitativa (la r de Pearson) y proporción. En todas las situaciones que se consideraron se trató con parámetros de poblaciones y hacen parte de un tema general de la estadística: los tests de hipótesis.

Situaciones problemáticas y calculadoras gráficas

En esta sección queremos hacer más explícito lo que vamos a entender por ‘situación problemática’ y el papel que van a jugar las calculadoras gráficas en dichas situaciones. Butts (1980) utiliza la expresión ‘situación problemática’ cuando clasifica los problemas que se proponen a los alumnos en cuatro tipos: ejercicios algorítmicos, problemas de aplicación, problemas de búsqueda y situaciones problemáticas. Menciona que los ejercicios que

se realizan a mano o con calculadoras rudimentarias son en general, de los tres primeros tipos. Por otra parte, en sesiones de clases donde se implemente el uso de recursos tecnológicos tales como paquetes estadísticos, hojas electrónicas e incluso calculadoras gráficas, se pueden plantear situaciones problemáticas de análisis de datos. En nuestro caso sólo vamos a considerar la utilización de calculadoras gráficas.

El apoyo tecnológico que nos aporta la implementación de calculadoras gráficas, puede traer ventajas que también podrían considerarse como desventajas a la hora de poner en juego contenidos específicos de la estadística alrededor de situaciones problemáticas. Por ejemplo, con el apoyo de calculadoras gráficas pueden plantearse actividades que podrían clasificarse como problemas de aplicación en el sentido de que el alumno debe de trasladar el problema verbal a una formulación simbólica, pero en lugar de resolverlo algorítmicamente lo puede hacer mediante programas implementados en la calculadora.

Las situaciones problemáticas de tipo estadístico que nosotros vamos a considerar, contemplan desde el punto de vista del uso de la calculadora gráfica situaciones intermedias entre dos extremos, uno, en el que los estudiantes tienen que centrar la mayor parte de la atención en los cálculos involucrados en los algoritmos de resolución de los problemas, y el otro, que se consigue con el uso de paquetes estadísticos, en el que el estudiante no deben desarrollar destrezas en los procesos algorítmicos de cálculo, y debe centrar la mayoría de la atención en la interpretación de los resultados.

Por otra parte, para el diseño de cada situación nos impusimos la restricción de que se debía aplicar durante dos sesiones de clase, de tal forma que en general, en la primera sesión se trabajara individualmente y en la segunda se trabajara en grupos de dos o tres estudiantes. Previamente a la aplicación de las situaciones problemáticas, los estudiantes debían recibir instrucción en el manejo de las utilidades de tipo estadístico que se tienen disponibles en la calculadora gráfica (Texas Instruments TI-85).

APROXIMACIÓN CONCEPTUAL DE LA PROPUESTA

La necesidad de diseñar situaciones problemáticas, llevarlas al salón de clase para su aplicación y evaluarlas nos obliga a asumir una posición con respecto a su planificación curricular. Para Rico (1997) hay diferentes niveles de organización de un currículo.

Los diferentes niveles de reflexión sobre el currículo de matemáticas han surgido al poner el énfasis sobre el currículo desde una consideración o planteamiento teórico determinado. así, cuando hemos asumido el currículo como un plan de acción para el profesor, se ha establecido un nivel microsociológico, desde el nivel de planificación para la actuación en el aula; cuando hemos considerado el cu-

currículo como una planificación para la administración educativa, el nivel de actuación es intermedio y está en el sistema educativo. Cuando se acepta el currículo como objeto de estudio, estamos en un nivel de reflexión académico y cuando atendemos a la aportación que deben hacer la matemáticas a los fines generales de la educación, nos hemos situado en una perspectiva teleológica. En cada uno de los niveles de reflexión mencionados el currículo se ha podido caracterizar mediante cuatro componentes que proporcionan el núcleo de conceptos más adecuados para organizar ese nivel. (Rico, 1997, p. 409)

La discusión que hace Rico acerca de los diferentes niveles de planificación del currículo se resume en la Tabla N° 2.

Tabla N° 2. Componentes del currículo según los niveles y dimensiones (tomado de (Rico, 1997, p. 409).

		Dimensiones del currículo			
		Cultural/ Conceptual	Cognitiva o de desarrollo	Ética o formativa	Social
Niveles	Planificación para los profesores	Contenidos	Objetivos	Metodología	Evaluación
	Sistema educativo	Conocimiento	Alumno	Profesor	Aula
	Disciplinas académicas	Epistemología e historia	Teorías de aprendizaje	Pedagogía	Sociología
	Teleológico o de finalidades	Fines culturales	Fines formativos	Fines políticos	Fines sociales

En primer lugar, nosotros queremos situarnos en el tercer nivel, es decir, en el que se mira el currículo como un objeto de estudio. Sin embargo, nuestra aproximación es modesta pues de ningún modo debemos pretendemos abordar todas las dimensiones expuestas ni tampoco profundizar en las consideradas con todo el detalle que se podría llegar a hacer. Sólo hemos mencionado la cita de Rico para delinear el tipo de trabajo que es posible realizar en cada dimensión. Nosotros sólo vamos a tener en cuenta tres de estas dimensiones. Una, es la dimensión cultural-conceptual, en la que se debe ir a la historia y a la epistemología como fuentes de consulta. Entonces, bajo esta dimensión, vamos en los capítulos 2 y 3 a hacer énfasis en el análisis del contenido estadístico bajo esta perspectiva. La segunda, es la dimensión cognitiva o del desarrollo, en la que se deben considerar las teorías acerca del aprendizaje y establecer nuestra posición en cuanto a lo que entendemos por aprendizaje y comprensión, esto lo vamos a exponer en los ca-

pítulos 4 y 5. Y la tercera dimensión, es la ética o formativa, en la que debemos considerar los diferentes aspectos relacionados con la pedagogía y la enseñanza. Vamos a considerar algunos de estos aspectos en los capítulos 6 y 7. La cuarta dimensión, que no fue considerada en este trabajo, considera el aspectos social y se apoya preferencialmente en estudios de sociología.

2. ELEMENTOS PARA EL ANÁLISIS DEL CONTENIDO

INTRODUCCIÓN

Realizar un análisis de la estructura del contenido no es simplemente detallar y explicitar los temas que se deben tratar; nosotros pensamos, que una mirada a los contenidos estadísticos debe contemplar no sólo la naturaleza del conocimiento estocástico y sus modos de representación, sino también a la manera como estos funcionan en relación con la resolución de problemas.

Con respecto a la naturaleza del conocimiento estocástico es muy diciente la argumentación que hace Moore (1992), en la que distingue entre matemáticas y estadística como disciplinas separadas. Para Moore la estadística es una ciencia matemática pero no es una rama de las matemáticas, emerge como una disciplina que tiene sus propios modos característicos de pensamiento, los cuales son más fundamentales que métodos o teorías matemáticas específicas. Por su parte, Gal y Garfield (1997) dicen:

En estadística, el contexto motiva los procedimientos y es la fuente de significado y la base para la interpretación de resultados de tales actividades. Moore (1990) señala que los datos deben ser vistos como números con un contexto. (p. 6)

En cuanto a lo que se refiere a la relación del conocimiento matemático y en particular del conocimiento estadístico con la resolución de problemas, es interesante tener en cuenta lo que dice Douady (1986). Ella habla del funcionamiento de los conceptos matemáticos y mira dicho funcionamiento bajo dos perspectivas: como útil y como objeto.

Un concepto es útil cuando focalizamos nuestro interés sobre el uso que se hace de él para resolver problemas [...] Por objeto, entendemos el objeto cultural que tiene lugar en un edificio más amplio que es el saber sabio en un momento dado, reconocido socialmente". (p. 9)

La cita de Douady, nos hace ver como el conocimiento tiene diferentes niveles de concreción. En el último nivel, el nivel más erudito, estaría el conocimiento como objeto cultural. Por otra parte, para mirar el conocimiento escolar se debe tener una perspectiva más amplia. El conocimiento no se da de la misma manera cuando se tiene que impartir en las instituciones educativas, ocurre en la adaptación del saber como objeto al conocimiento escolar. Esto es lo que los franceses denominan *transposición didáctica*., Brousseau (1993) dice:

El conocimiento matemático escolar enmascara el verdadero funcionamiento de la ciencia, imposible de comunicar y describir fielmente desde el exterior, para poner en su lugar una génesis ficticia. Para facilitar la enseñanza, aísla ciertas nociones y propiedades del tejido de actividades en donde han tomado su origen, su sentido, su motivación y su empleo. Ella los transpone en el contexto escolar. (p. 3)

Como profesores-investigadores debemos ser conscientes que la transposición didáctica ocurre inevitablemente cuando se lleva el conocimiento estadístico a la institución escolar. Ello nos obliga a mirar dicho conocimiento de una manera más reflexiva y a fijar un marco conceptual para concretar la manera como debemos realizar dicha mirada.

En primer lugar, para realizar la descripción del conocimiento estadístico nos acogemos a la propuesta de Rico (1995a). Allí se plantea la organización del conocimiento matemático bajo dos perspectivas, la conceptual y la procedimental. Encontramos en este modelo de organización, una manera de pensar acerca del contenido que nos sugiere una forma de proceder metodológicamente para hacer una descripción más elaborada del contenido. En segundo lugar, consideramos como parte del análisis de contenido, el uso del lenguaje y en particular, de cuatro formas de representación externa del conocimiento: la simbólica, la gráfica, la tabular y la verbal. Finalmente y en tercer lugar, también vamos a considerar algunos aspectos de la resolución de problemas de tests de hipótesis como parte del análisis del conocimiento estocástico.

CONOCIMIENTO CONCEPTUAL Y PROCEDIMENTAL

La relación entre conocimiento conceptual y destreza procedimental es uno de los aspectos claves no sólo para el aprendizaje de las matemáticas y la estadística, sino también para el aprendizaje en general. Hiebert y Lefevre (1987), mencionan cómo en diferentes áreas del conocimiento estas diferencias parecen darse en la terminología y en el énfasis utilizado según la disciplina científica.

[...] las diferencias son principalmente de énfasis más que de especie. Por ejemplo Piaget (1978) distingue entre comprensión conceptual y acción con éxito: Tulving (1983), distingue entre memoria semántica y memoria episódica; Anderson (1983) distingue entre conocimiento declarativo y conocimiento procedimental. Paralelas distinciones se hacen en las teorías filosóficas del conocimiento. Por ejemplo Scheffler (1965) distingue entre el uso proposicional del 'conocer que' y el uso procedimental de 'conocer como'. (p. 1)

Un recuento histórico acerca del debate entre lo conceptual y lo procedimental llevaría a reconocer, prescindiendo de las etiquetas, a la distinción entre destreza y comprensión (Vallecillos, 1996). Para fijar una posición con respecto al conocimiento conceptual y procedimental adherimos a la posición que Hiebert y Lefevre (1987):

A pesar de que las recientes orientaciones sobre el tema del conocimiento conceptual y procedimental proporcionan nuevas y significativas ideas sobre el aprendizaje de las matemáticas y sus aplicaciones, las relaciones entre estas formas de conocimiento no son aún bien comprendidas. Un primera razón para la naturaleza espinosa del problema es que los propios tipos de conocimiento son difíciles de definir. El núcleo de cada uno de ellos es fácil de describir, pero los límites son difíciles de delimitar. Nuestra posición es que la distinción entre conocimiento conceptual y procedimental es útil para pensar acerca del aprendizaje de las matemáticas [...] No creemos, sin embargo, que la distinción proporcione un esquema de clasificación en el cual toda clase de conocimiento pueda ser tipificado [...] No obstante creemos que es posible distinguir entre los dos tipos de conocimiento y que tal distinción proporciona una manera de interpretar los procesos de aprendizaje que nos ayuda a comprender mejor los fallos y los éxitos de los estudiantes. (p. 3)

Siguiendo a estos autores y a Rico (1995a), presentamos a continuación las principales características del conocimiento conceptual y procedimental.

Conocimiento conceptual

Este conocimiento se puede entender como una red de unidades de información interrelacionadas, donde se considera tan importante la relación entre las distintas unidades, como las piezas discretas de información. Una unidad de información no se considera como tal, hasta tanto el sujeto no la reconozca como relacionada con otra unidad.

Se supone que el conocimiento conceptual se construye elaborando nuevas relaciones entre unidades de información previamente almacenadas o a través de la integración de nuevas unidades de información con las redes existentes. Rico (1995a) distingue tres niveles de conocimientos en el campo conceptual: los hechos, que son unidades de información y sirven como registros de acontecimientos; los conceptos propiamente tales, que describen una regularidad o relación de un grupo de hechos, suelen admitir un modelo o representación y se designan con un signo o un símbolo; y las estructuras conceptuales, que sirven para unir conceptos o para sugerir formas de relación entre conceptos constituyendo, a veces, conceptos de orden superior, ya que pueden establecer algún orden o relación entre conceptos no inclusivos.

Conocimiento procedimental

Una de las características principales del conocimiento procedimental es su naturaleza secuencial y estructurada. En éste se integran los símbolos y la sintaxis del lenguaje matemático con las reglas y procedimientos que permiten resolver un problema.

Los procedimientos son aquellas formas de actuación o ejecución de tareas matemáticas. Rico (1995a), también distingue tres niveles diferentes en el campo de los procedimientos: las destrezas, que permiten transformar una expresión simbólica desde una forma dada hasta otra forma, y para lo cual hay que ejecutar una secuencia de reglas sobre manipulación de símbolos; los razonamientos que se presentan al procesar relaciones entre conceptos, y permiten establecer relaciones de inferencia entre los mismos; y las estrategias, que se ejecutan sobre representaciones de conceptos y relaciones, operan dentro de una estructura conceptual y suponen cualquier tipo de procedimiento que pueda ejecutarse, teniendo en cuenta las relaciones y conceptos implicados.

El conocimiento de tipo procedimental se puede adquirir de una manera mecánica, sin significado. Cuando los conocimientos son aprendidos de una manera significativa se considera que están ligados con el conocimiento conceptual, sin embargo, es frecuente encontrar situaciones donde es difícil de separar completamente los dos tipos de conocimiento.

SISTEMAS DE REPRESENTACIÓN

En la práctica de las matemáticas y en particular en la de la estadística, la representación es una actividad usual. Los modos de expresión de conceptos y de leyes matemáticas y estocásticas vienen dados mediante representaciones externas. Hiebert y Carpenter (1991) sostienen que “para pensar sobre ideas matemáticas y comunicarlas necesitamos representarlas de algún modo.” La comunicación requiere de representaciones externas, que toman la forma de lenguaje oral, símbolos escritos, dibujos u objetos físicos. Sin embargo, “[...] para pensar sobre ideas matemáticas necesitamos representarlas internamente, de manera que permita a la mente operar sobre ellas.” La relación entre las representaciones internas y externas es un tema sobre el cual nos vamos a referir con más detalle en el capítulo sobre análisis del aprendizaje.

En este apartado solamente queremos explicitar algunas formas concretas de representación externa. Entre ellas aparecen referidas con más frecuencia la simbólica y la gráfica. Sin embargo, consideramos importante hacer una breve referencia a cuatro formas de representación externa. En particular, usamos la propuesta de Janvier (1987) que, aunque es utilizada para explicar el manejo del concepto de variable en la resolución de problema, también se puede aplicar en otros contextos. Janvier (1987) además

hace explícito en una tabla el tipo de procesos de traducción que ocurren normalmente. La siguiente es la tabla presentada por él.

Tabla N° 3.

a De	Situaciones (descripciones verbales)	Tablas	Gráficas	Fórmulas
Situaciones (descripciones verbales)		Medir	Dibujar	Modelar
Tablas	Leer		Marcar puntos	Ajustar
Gráficas	Interpretar	Terminación de una lectura		Ajustar curvas
Fórmulas	Reconocer parámetros	Calcular	Dibujar	

Vale la pena hacer algunos comentarios acerca de las cuestiones que se pueden entrever en esta tabla. En primer lugar, la forma como fue llenada la tabla. No hay, en general, una manera única de llenar dicha tabla. Esto puede depender en gran parte del contexto en el que se está trabajando. En segundo lugar, a pesar de las limitaciones que se le pueda encontrar, la tabla es muy útil para dar una idea general de las preguntas que pueden surgir en los procesos de resolución de problemas, ayuda de una manera simple a ilustrar una variedad de aspectos relacionados con el manejo de conceptos y procedimientos y permite ver toda la complejidad que hay detrás de un análisis de esta naturaleza. Finalmente, Janvier también señala que aunque la diagonal no se haya llenado, también allí se dan procesos de cambio entre un mismo tipo de representación, él propone llamarlo transposición.

RESOLUCIÓN DE PROBLEMAS EN TESTS DE HIPÓTESIS

La resolución de problemas usualmente requiere la aplicación tanto del conocimiento conceptual como del procedimental. Por ello se tiene en los problemas uno de los medios más apropiados para el estudio de los dos tipos de conocimiento. Además, si se tiene en cuenta el uso que se hace del contraste de hipótesis para la resolución de problemas, se encuentra en este tipo de actividad uno de los instrumentos más útiles para valorar la competencia conseguida por nuestros estudiantes.

Chevarney et al. (1977) y también Hiebert y Wearne (1987) consideran la resolución de problemas como un proceso donde se pueden identificar diferentes momentos o estados en cada uno de los cuales se reconocen las relaciones entre ambos tipos de conocimiento como especialmente productivas. Estos autores, independientemente del contenido del problema plantean tres fases en la resolución de la siguiente manera:

- Fase I. Comprensión del problema: que incluye comprender el enunciado, identificar el tipo de problema y los conceptos estadísticos implicados en el enunciado y reconocer los datos conocidos y desconocidos.
- Fase II. Planificación y ejecución: que considera la selección de las técnicas estadísticas adecuadas para resolver el problema y el uso correcto de las mismas para obtener la solución pedida.
- Fase III. Evaluación e interpretación: que consiste en la verificación de la plausibilidad de la solución y en la interpretación de la misma en términos estadísticos y del contexto del problema.

De otro lado, en Chatfield (1988), encontramos una exposición muy elaborada sobre las fases y características que se deben tener en cuenta al hacer el planteamiento y llevar a cabo la solución de un problema de tipo estadístico. Aunque su propuesta es general y no se restringe a los tests de hipótesis, allí se pueden encontrar algunos ejemplos de problemas que motivan el uso de tests de hipótesis cuyas soluciones son comentadas.

3. ANÁLISIS DEL CONTENIDO EN LOS TESTS DE HIPÓTESIS

INTRODUCCIÓN

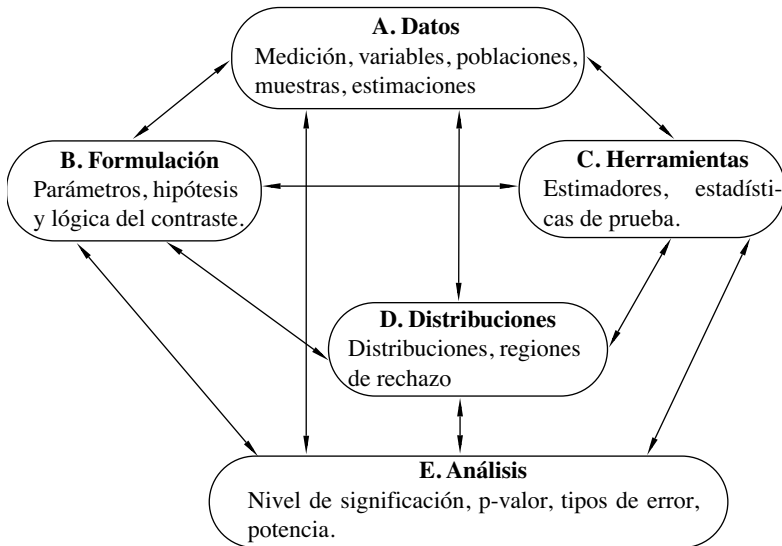
En esta parte vamos a hacer una descripción de los conocimientos que se relacionan con un test de hipótesis. Para describir de una manera organizada dichos conocimientos, vamos, en primer lugar, a separar en cinco unidades los diferentes conceptos y procedimientos asociados al test de hipótesis. Luego y con base en esta división vamos a presentar un diagrama de estructura conceptual, en el cual se pueden considerar al menos diez conjuntos de conexiones entre los diferentes conceptos y procedimientos.

UNIDADES DE ANÁLISIS

Las cinco unidades de análisis que vamos a considerar son las siguientes:

- A. Datos
- B. Formulación
- C. Herramientas
- D. Distribuciones
- E. Análisis (Interpretación de resultados)

El gráfico que se presenta a continuación muestra de manera esquemática las unidades que se consideran. En el interior de la caja que representa a cada unidad se han escrito los nombres de algunos conceptos que consideramos hacen parte de ella.

*Figura 1.*

A continuación haremos algunos comentarios acerca de cada unidad y de sus elementos. Debido a lo extensivo que se puede hacer la exposición, se ha decidido hablar sólo acerca de los conceptos y de sus relaciones y no de los procedimientos asociados. Más adelante presentaremos en la Tabla N° 4, una lista detallada de conceptos y procedimientos asociados y en la Tabla N° 5, una lista de las principales relaciones identificadas entre conceptos.

A. Datos. Esta unidad considera conceptos tales como: medición, variable, población, muestra, nivel de significación y estimaciones. Por ejemplo, cuando hablamos de medición usualmente nos referimos a las escalas; términos como nominal, ordinal, intervalo y razón, tienen un significado concreto que debe ser claro como parte de los datos de un problema.

Podemos referirnos a una variable como una característica que puede adoptar distintos valores, es una cualidad o aspecto en el cual difieren un conjunto de individuos u objetos. Bisquerra (1989) refiriéndose al concepto de variable en estadística dice que las variables se pueden definir de forma constitutiva o de forma operativa; en el primer caso, están relacionadas con el contexto de la situación problemática y se identifican con el constructo o esencia del fenómeno considerado; en la definición operativa, más bien se describen las operaciones a realizar para medirla o manipularla. Bisquerra también presenta cuatro criterios a partir de los cuales se pueden clasificar las variables: teórico-explicativo, metodológico, de medición y de control. Si miramos las variables en relación con el concepto de medición, se pue-

den clasificar como cualitativas (categóricas dicótomas o politómicas) y en cuantitativas (discretas y continuas). Por otra parte, si miramos el problema estadístico desde el punto de vista del diseño metodológico, entran en juego las nociones de variables independientes, dependientes y extrañas o desde el punto de vista de control, de las variables aleatorias y las variables controladas.

También las nociones de población y muestra son importantes. La población de estudio normalmente es entendida como el conjunto de todos los individuos u objetos que se desea estudiar. Sin embargo, lo que se analiza en la práctica son las poblaciones de datos que se asocian a una población de estudio. En el caso de las muestras, también debe distinguirse entre la muestra de estudio, como subconjunto de una población de estudio y las muestras de datos que se obtienen de cada elemento de la muestra de estudio.

Con respecto al concepto de nivel de significación, primero que todo se debe advertir que es un concepto bastante complejo. En esta unidad simplemente lo queremos mirar como “un dato”, esto quiere decir que consideramos que sólo adquiere importancia dentro de la resolución de problemas cuando se quiere determinar una región crítica para una decisión acerca de las hipótesis o cuando se le compara con un p-valor. Por otra parte, la elección de un nivel de significación normalmente es una elección de carácter subjetivo y que está más ligada al contexto del problema específico que se esté tratando.

B. Formulación. Esta unidad considera el planteamiento y la formulación del problema estadístico en términos de la identificación y discriminación de dos conceptos que están relacionados: el de parámetro y el de hipótesis. También haremos referencia a la lógica que hay detrás del test de hipótesis.

En relación con las hipótesis, normalmente se establece la diferencia entre las hipótesis de investigación y las hipótesis estadísticas. En general las hipótesis estadísticas (exceptuando el caso de las hipótesis de tipo no paramétrico), están definidas en términos de los parámetros y éstos a su vez en términos de algunas de las variables consideradas en el problema.

La formulación de las hipótesis estadísticas, la hipótesis nula y la alternativa, son el resultado de un proceso de formalización de la hipótesis de investigación. Usualmente, la hipótesis de investigación debe traducirse a la forma de hipótesis estadísticas cuando se desea utilizar como herramienta de análisis en un test de hipótesis. Si bien es cierto, que también es posible en muchos casos se pueden utilizar los intervalos de confianza para tomar una decisión acerca de un test de hipótesis, los intervalos de confianza son una herramienta que se utiliza de manera más natural, cuando no se tienen hipótesis de investigación o afirmaciones a priori acerca de los parámetros.

Para formular una hipótesis en forma estadística debe ser claro cuál o cuáles son las variables de medición, qué parámetro se va a estudiar y en qué forma se van a contrastar. La formulación de las hipótesis nula y alternativa dependen de lo que se plantee como hipótesis de investigación. En este sentido, la direccionalidad —unilateral, o bilateral— y el rango de valores a los que se refieren las hipótesis —simples o compuesta— son dos aspectos que normalmente se deben considerar.

Con respecto a la lógica del razonamiento en los contrastes de hipótesis hay varias cuestiones interesantes que vale la pena mencionar. En primer lugar debe ser claro la diferencia entre el tipo de prueba que proporciona un contraste y el que proporciona un razonamiento deductivo. Detrás del razonamiento estadístico de los contrastes de hipótesis subyace una lógica que aparentemente podría compararse con el razonamiento que se sigue al realizar una prueba por contradicción¹. En realidad el razonamiento estadístico que se debe hacer cuando se trabaja con los contrastes de hipótesis es mucho más complejo, que lo que podría argumentarse haciendo un parangón con la prueba por contradicción.

Una revisión histórica acerca de este tema, da cuenta de tres escuelas diferentes sobre las cuales ha evolucionado la lógica sobre la que se apoya este tipo de razonamiento estadístico. En primer lugar se puede mencionar la escuela de Fisher que se enfoca exclusivamente sobre la hipótesis nula. Para Fisher, una vez planteada la hipótesis nula, las observaciones sólo permiten rechazar la hipótesis pero nunca confirmarla, el error de tipo II no tiene sentido en este tipo de razonamiento, pues no se considera el error que se comete al aceptar una hipótesis nula, supuesto que sea falsa. En cierta forma los tests de significación de Fisher presentan una analogía lógica con la epistemología de Popper de rechazo de la hipótesis. (Popper, 1967)

Por contraste, en la escuela de Neyman-Pearson sí se establece realmente un contraste entre dos hipótesis. Mientras que para Fisher los tests de significación eran propuestos como modelos de razonamiento inductivo, el programa de Neyman-Pearson se concibe como una regla de conducta inductiva, por lo que esta teoría pertenece más bien a la teoría de la decisión que a la lógica inductiva. Con la propuesta de Neyman-Pearson se nos permite obtener probabilidades de las muestras particulares, a partir de una hipótesis dada que se supone cierta y no a la inversa.

La tercera escuela, se fundamenta en el teorema de Bayes, el cual constituye una supuesta solución al cálculo inductivo de la probabilidad de una hipótesis, dado que permite calcular una probabilidad a posteriori, en fun-

1. En una demostración por contradicción donde se quiere demostrar una hipótesis “p” se asume que es cierto “no p” y se argumenta para llegar a una contradicción, es decir, para llegar a que para cierta premisa “q”, es cierto “q” y “no q”; como consecuencia de ello se concluye que no puede ser “no p”, luego tiene que ser “p”.

ción de la información obtenida. El problema con esta posición es que el conocimiento de la probabilidad a priori de las causas, implica una gran dificultad metodológica, por ello mucho autores, entre ellos Fisher, rechazan el argumento bayesiano. Sin embargo, según Hacking (1965), la teoría de los tests de significación de Fisher carece de fuerza porque siempre nos llevaría al rechazo de cualquier hipótesis. Para ello, basta con tomar un tamaño de muestra que sea suficientemente grande, con lo que se asegura que un pequeño efecto sea estadísticamente significativo. Hacking también critica la teoría de Neyman-Pearson por sus características de teoría inferencial inversa, es decir, que sólo vale antes de observar los datos, pero no después. Hay una gran diferencia para el investigador en el significado que tiene el nivel de significación antes y después de tener conocimiento del resultado de un experimento. Por ejemplo, antes de conocer el resultado, un nivel de significación pequeño, proporciona una buena oportunidad para apostar al rechazo erróneo de una hipótesis nula. Sin embargo, una vez conocido el resultado, no todos los tests con el mismo nivel de significación proporcionan apuestas igualmente buenas.

C. Herramientas. En esta unidad de análisis se consideran los estimadores, las estimaciones y los tests estadísticos que se asocian con los tests de hipótesis. Las estadísticas de prueba que se van a mencionar, con excepción del caso de experimentos binomiales, son tests de hipótesis de tipo paramétrico. En particular, se van a considerar estadísticas de prueba para la diferencia de medias muestrales con base en dos muestras independientes o relacionadas; para inferir acerca del parámetro de proporción se va a considerar el test no paramétrico basado en la distribución binomial y también el test paramétrico, basado en la aproximación normal de la distribución binomial; finalmente, se van a considerar algunas estadísticas de prueba para inferir acerca del coeficiente correlación poblacional de tipo lineal (correlación de Pearson).

Uno de los elementos esenciales que se trabajan cuando se considera una estadística de prueba de tipo paramétrico, es el de error estándar, que es precisamente la desviación estándar de la distribución muestral del estimador del parámetro que se está estimando.

Por otra parte, en un nivel superior de conocimiento, la elección de una herramienta de análisis estadístico exige no sólo el conocer las diferentes alternativas que hay disponibles sino también los supuestos en que se basa su aplicación y las posibles dificultades técnicas para su ejecución. En general cualquier estadística de prueba de tipo paramétrico tiene su contraparte en el área no paramétrica, pero en ésta última, varían, en general, los supuestos de un test.

D. Distribuciones. En cuanto a los modelos de distribución, la teoría de la probabilidad nos ofrece diversos modelos de distribuciones cuya utiliza-

ción en determinados tipos de problemas han probado su eficacia. En el caso de las situaciones problemáticas que nos ocupan las distribuciones normal, t-student y la binomial van a jugar un papel central. El modelo binomial surge de una replicación de ensayos de variables de Bernoulli donde en cada ensayo nos fijamos solamente en la ocurrencia de un “éxito” o un “fracaso” con una probabilidad p de “éxito” que permanece constante a lo largo de los ensayos. Por otra parte, la importancia del modelo normal se debe a varias razones entre las cuales cabe destacar:

- Muchos teoremas de límites muestran que en condiciones muy generales la suma de un gran número de variables aleatorias tengan una distribución asintóticamente normal.
- Debido a estos teoremas, la distribución de muestreo de ciertos estadísticos, como la media o la proporción, tienen distribuciones aproximadamente normales para tamaños de muestras suficientemente grandes.

Por otra parte, conceptos tales como el de distribución muestral, distribución de la población, distribución de la muestra, distribución teórica y distribución empírica, son trascendentales para establecer distinciones conceptuales.

Cuando en los tests de hipótesis se encuentran valores críticos para establecer las regiones de rechazo y de no rechazo, estos valores se determinan con base en modelos de distribución teóricos que aproximan el comportamiento de una distribución muestral.

E. Análisis. En esta unidad conceptual consideramos los conceptos que, a nivel de los tests de hipótesis, son relevantes para el análisis y la interpretación de los resultados de una estadística de prueba. Entre otros, se pueden nombrar los tipos de errores relacionados con el proceso de decisión, conocidos como errores de tipo I y II, los conceptos de nivel de significación y de p-valor y el concepto de potencia de una prueba estadística también asociado con los tipos de error I y II.

Uno de los conceptos que más le ha llamado la atención a los investigadores en el campo experimental y también en la didáctica de la estadística es el de nivel de significación. Por ejemplo, se encuentran antiguas referencias tales como la Morrison y Henkel (1970) y referencias recientes tales como la de Falk y Greenbaum (1995).

Los tests de hipótesis siguen una tradición en la que tanto las hipótesis como el nivel de significación α son fijados antes de tomar las observaciones muestrales en el modelo de investigación experimental. Es un enfoque en el que los datos se recogen con el propósito de probar un modelo teórico formulado con anterioridad. Sin embargo, a partir de las ideas de Tukey (1977), quien inicia la corriente del análisis exploratorio de datos y se desa-

rolla gracias a las posibilidades del análisis multivariante, las hipótesis son sugeridas a partir de los datos. Grass (1992) argumenta que esto supone una ruptura epistemológica con la inferencia clásica que tiene que ver con “los objetivos, los medios técnicos para lograrlos, los datos tratados, los sujetos del análisis, los modos de restitución de la información y, el método (de los datos a los modelos y no a la inversa).” (p. 64)

CONCEPTOS Y PROCEDIMIENTOS

A continuación se presenta la Tabla N° 4 en la que al lado de cada unidad conceptual, se presenta una lista de los principales conceptos asociados con la unidad, algunos de los posibles procedimientos relacionados y también la fase que se le puede asociar en la resolución de problemas.

Tabla N° 4. Conceptos y procedimientos asociados con las unidades conceptuales del contraste de hipótesis y con las fases de resolución de problemas.

Conceptos asociados y fase involucrada	Procedimientos asociados
Variable (Fase I)	A1. Identificar la(s) variable(s) de estudio. A2. Distinguir entre una variable independiente y una dependiente. A3. Distinguir entre una variable continua y una discreta
Muestra (Fase I)	A4. Identificar el número de muestras que se consideran. A5. Cuando en el problema se trabaja con dos muestras, identificar si las muestras son independientes o relacionadas. A6. Establecer un criterio para determinar si los tamaños de las muestras son “grandes” o “pequeñas”. A7. Identificar el tamaño de la muestra.
Población (Fase I)	A8. Establecer cuál es la población de estudio y cuál la de datos. A9. Distinguir la población de la muestra.
Nivel de significación (Fase I)	A10. Identificar el nivel de significación dado en un problema. A11. Determinar un nivel de significación para resolver el problema. A12. Trasladar un valor porcentual a uno proporcional o viceversa.

Tabla N° 4. Conceptos y procedimientos asociados con las unidades conceptuales del contraste de hipótesis y con las fases de resolución de problemas.

Conceptos asociados y fase involucrada	Procedimientos asociados
Medidas de resumen estadístico (Fase I)	A13. Identificar y distinguir los promedios, las proporciones, las desviaciones y/o las correlaciones poblacionales de las muestrales. A14. Interpretar los valores de los estadísticos de resumen.
Parámetros (Fase I)	B1. Identificar o reconocer el o los parámetros de estudio. B2. Distinguir un parámetro de su estimador o del valor de una estimación.
Hipótesis de Investigación (Fase I)	B3. Identificar la hipótesis de investigación del problema. B4. Establecer la relación entre la hipótesis de investigación del problema y el parámetro de estudio.
Hipótesis estadísticas (Fase I)	B5. Formular las hipótesis refiriéndose a parámetros de las poblaciones. B6. Establecer en las hipótesis estadísticas los valores posibles de los parámetros.
Lógica del contraste (Fase I)	B7. Establecer el contraste de hipótesis como un problema de decisión entre dos hipótesis estadísticas. B8. Elegir como hipótesis nula la contraria a la que se desea confirmar. B9. Establecer antes de observar los resultados del contraste, las hipótesis.
Estimador (Fase II)	C1. Calcular valores de los estimadores. C2. Ubicar valores de las estimaciones en un gráfico.
Varianza del estimador (Fase II)	C3. Identificar el error estándar que se asocia con el estimador. C4. Calcular el valor del error estándar.
Estadístico de prueba (Fase II)	C5. Distinguir entre el estimador y el estadístico de prueba. C6. Calcular el valor del estadístico de prueba. C7. Representar en una gráfica el valor de la estadística de prueba.
Distribución estimador (Fase II)	D1. Identificar una distribución de probabilidad asociada al estimador. D2. Identificar los parámetros asociados a la distribución. D3. Representar de manera gráfica la distribución del estimador.

Tabla N° 4. Conceptos y procedimientos asociados con las unidades conceptuales del contraste de hipótesis y con las fases de resolución de problemas.

Conceptos asociados y fase involucrada	Procedimientos asociados
Distribución estadístico de prueba (Fase II)	<p>D4. Identificar una distribución de probabilidad asociada al estadístico de prueba.</p> <p>D5. Identificar los parámetros de la distribución.</p> <p>D6. Representar de manera gráfica la distribución del estadístico.</p> <p>D7. Consultar la tabla de distribución del estadístico para hallar un valor de la distribución dado un nivel de significación.</p> <p>D8. Determinar el p-valor de un valor de la estadística de prueba.</p> <p>D9. Establecer criterios de decisión.</p> <p>D10. Representar gráficamente las regiones de rechazo y de no rechazo de la hipótesis nula.</p> <p>D11. Representar de manera gráfica el p-valor.</p>
Distribución muestral (Fase II)	<p>D12. Distinguir entre la distribución muestral y el modelo de distribución teórico que aproxima dicha distribución.</p> <p>D13. Revisar los supuestos que validan la aproximación escogida.</p>
p-valor (Fase III)	<p>E1. Interpretar un valor “pequeño” del p-valor como favoreciendo el rechazo de la hipótesis nula.</p> <p>E2. Interpretar un valor “grande” del p-valor como favoreciendo el no rechazo de la hipótesis nula.</p>
Nivel de significación (Fase III)	<p>E3. Comparar el nivel de significación y el del p-valor para tomar una decisión.</p> <p>E4. Comparar el valor crítico asociado al nivel de significación con el valor calculado del estadístico de prueba.</p>
Tipos de error (Fase III)	<p>E5. Identificar el tipo de error que se puede cometer.</p> <p>E6. Evaluar las consecuencias de cometer un error de tipo I.</p> <p>E7. Determinar la magnitud del error de tipo I.</p> <p>E8. Evaluar las consecuencias de cometer un error de tipo II.</p> <p>E9. Determinar la magnitud del error de tipo II.</p>
Potencia (Fase III)	<p>E10. Calcular la potencia de la prueba para una alternativa dada.</p> <p>E11. Construir una curva de potencia.</p> <p>E12. Determinar la potencia de la prueba.</p>

También se establecieron numerosas relaciones entre los conceptos mencionados en la tabla anterior. Algunas de estas relaciones las hemos tomado de Vallecillos (1996), ella se refiere a estas relaciones como elementos de significado asociados a un contraste de hipótesis. En la Tabla N° 5 que presentamos mas adelante se enumeran estas relaciones y en la última columna se hace referencia a las conexiones que se pueden establecer entre las diferentes unidades de análisis con relación a la interrelación enunciada.

Tabla N° 5. Relaciones entre algunos de los conceptos estadísticos enunciados y tipo de conexión asociada entre las unidades conceptuales de los tests de hipótesis.

N°	Interrelaciones	Tipo
1	En la formulación del contraste de hipótesis se está ante un problema de decisión entre dos hipótesis de las cuales no sabemos cuál es verdadera.	B y E
2	Hay que diferenciar entre hipótesis nula e hipótesis alternativa. Cada una de ellas juega un papel diferente en el contraste.	B y C
3	Hay que diferenciar entre hipótesis unilaterales y bilaterales. El procedimiento de contraste se ve afectado por esta distinción.	B y C
4	Existe la posibilidad de cometer un error de tipo I: rechazar una hipótesis nula verdadera.	B y E
5	El riesgo del error de tipo I viene medido por el nivel de significación.	A y E
6	Existe la posibilidad de cometer un error de tipo II: aceptar una hipótesis nula falsa.	B y E
7	Las hipótesis paramétricas se refieren a valores de los parámetros de las poblaciones	A y B
8	No se puede saber con seguridad si se comete uno de los dos tipos de error, ni siquiera una vez tomada la decisión.	E
9	Es diferente el tipo de prueba que proporciona un contraste y el que proporciona un razonamiento deductivo.	E
10	El estadístico de prueba es una estandarización del estimador	C
11	El nivel de significación no es una probabilidad inductiva de la hipótesis nula	A y B
12	El estadístico es una función de los valores de la muestra, es una variable aleatoria	A y C
13	El hecho de que el estadístico es una función de los valores de la muestra hace que el valor del estadístico en la muestra particular se tome como función de decisión.	A y C
14	La distribución del estadístico depende del valor asociado a parámetros de la población.	D y B

Tabla N° 5. Relaciones entre algunos de los conceptos estadísticos enunciados y tipo de conexión asociada entre las unidades conceptuales de los tests de hipótesis.

N°	Interrelaciones	Tipo
15	El nivel de significación es el área determinada por el valor crítico en la función de densidad de la distribución del estadístico en el muestreo, bajo la hipótesis nula.	D y B
16	Un resultado estadísticamente significativo es un resultado cuya probabilidad de ocurrencia, en caso de ser cierta la hipótesis nula es menor que el nivel de significación.	E
17	La significación estadística no implica necesariamente la significación (relevancia) desde el punto de vista práctico.	E
18	Un contraste de hipótesis no determina la probabilidad de certeza de H_0 .	B y E
19	Un contraste de hipótesis no determina la probabilidad a posteriori de una hipótesis en función de los datos.	A-B y E
20	El parámetro es un valor constante desconocido en la población	A y B
21	La distribución de un estadístico suele depender del tamaño de la muestra.	A y D
22	Los resultados de un análisis de significación dependen del tipo de distribución utilizada en la prueba	D y E
23	El nivel de significación no se determina por la teoría estadística, es fijado por el investigador	E y A
24	La varianza muestral de un estadístico decrece en función del tamaño de la muestra	A y C
25	El nivel de significación, junto con las hipótesis nula y alternativa y el estadístico de prueba determinan las regiones crítica y de aceptación en un contraste.	AB y CD
26	El p-valor del resultado de un estadístico de prueba depende de los datos de la muestra, de la formulación de las hipótesis y de la distribución del estadístico	AB y CE
27	El nivel de significación proporciona un criterio de decisión probabilístico, no un criterio matemático certero e incuestionable	A y E
28	Un misma hipótesis nula y un mismo nivel de significación pueden determinar diferentes regiones crítica y de aceptación dependiendo de la lateralidad de la prueba	BD y E
29	Un misma hipótesis nula y un mismo nivel de significación pueden determinar diferentes regiones crítica y de aceptación dependiendo del tamaño de la muestra	AD y E

ANÁLISIS EN TÉRMINOS DE REPRESENTACIONES EXTERNAS

En primer lugar vamos a mencionar cuando consideramos que estamos utilizando uno u otro tipo de representación.

Representación verbal. Consideramos que estamos trabajando en este tipo de sistema de representación cuando nos referimos a contenidos, conceptos o situaciones de la estadística que vienen descritas en forma verbal o escrita en el lenguaje natural.

Representación gráfica. Estamos trabajando con la representación gráfica cuando utilizamos dibujos o diagramas para representar conceptos o procesos estadísticos. La estadística está inundada de formas gráficas de representación. Los histogramas, los gráficos de dispersión, los gráficos de distribuciones, los box-plot, etc.

Representación simbólica. Para referirnos a nociones o a conceptos de estadística, los símbolos permiten gran flexibilidad y poder de síntesis. Ejemplos de este tipo de representaciones son las expresiones para representar un promedio muestral o una desviación estándar muestral o las letras griegas μ y σ , que representan los respectivos parámetros de la población asociados a la media y la desviación estándar.

Representación tabular. Se utiliza de manera preponderante para representar conjuntos de datos y de distribuciones. Por ejemplo, las tablas de frecuencia, las tablas de contingencia, las tablas de los modelos de distribución normal, binomial, etc.

En general cada concepto estadístico se puede analizar en términos de sus diversas formas de representación. En este sentido, es imposible hacer un análisis exhaustivo en el que se consideren todos y cada uno de los conceptos que aparecen en la Tabla N° 4. Por ello, en lo que se refiere al análisis en términos de las formas de representación, nos vamos a centrar más adelante, sólo en dos conceptos particulares: el nivel de significación y el p-valor.

4. ELEMENTOS PARA EL ANÁLISIS DEL APRENDIZAJE

INTRODUCCIÓN

En esta parte vamos a exponer algunos de los elementos conceptuales que se tendrán en cuenta para el análisis del aprendizaje estadístico de los estudiantes. Aunque desde el punto de vista cognitivo vamos a considerar algunos aspectos generales que se enfocan en el conocimiento matemático, queremos, bajo esta perspectiva, analizar la comprensión de los estudiantes de los temas estadísticos que nos interesan. En la primera sección nos vamos a referir al tema de la comprensión; en la segunda, a las nociones de concepto y concepción; en la siguiente tocaremos el tema de los errores, las dificultades y los obstáculos del conocimiento y en la última a los sistemas de representación.

COMPRENSIÓN

Sierpinska (1994), considera la complejidad de lo que significa comprender un concepto matemático. Ella plantea y desarrolla una respuesta a la pregunta “¿qué significa comprender una u otra noción en matemáticas?”. En el campo estadístico uno puede preguntarse por ejemplo, ¿qué significa comprender la noción de dispersión? La noción de dispersión es “un concepto elemental pero con frecuencia un concepto técnico desconcertante” (Mason, 1996, p. 228). Se puede saber que la dispersión se puede medir a través de la desviación estándar, también se puede saber como calcularla a través de la fórmula incluso se pueden saber cosas acerca de la desviación

$$\sqrt{\sum \frac{\langle x - \bar{x} \rangle^2}{n}}$$

estándar tales como que es diferenciable y que el cuadrado elimina la posibilidad de obtener una expresión negativa, y aún así, es posible que no se comprenda la noción de dispersión, pues “falta la conexión concepto-imagen que me lleve a la comprensión” (Mason, 1996, p. 229). La propuesta de Sierpinska es considerar la comprensión como un *proceso mental*. Proceso mental en el que se describe la evolución de diversos *estados de comprensión*. Gómez (1996) describe de manera sucinta la manera como Sierpinska concibe la comprensión matemática:

Un estado de comprensión con respecto, por ejemplo, a un concepto matemático, es una “forma de ver las cosas” con respecto a aquello que atañe al concepto. Esta forma de ver las cosas [...] es producto

de las experiencias que el sujeto ha vivido con respecto a este concepto, [...] es el “modelo” mental que el sujeto ha construido y con el cual él se ha sentido “cómodo” cuando tiene que actuar en situaciones matemáticas que involucran al concepto.

Sierpinska utiliza la noción de *acto de comprensión*. Ella propone “concebir la comprensión como un acto (de asir el significado) y no como un proceso o manera de conocer, [...] propongo, entonces, considerar la comprensión como un acto, pero un acto inmerso en un proceso de interpretación, siendo esta interpretación el desarrollo de una dialéctica entre adivinaciones más y más elaboradas y validaciones de estas adivinaciones”.

CONCEPTO Y CONCEPCIÓN

La escuela francesa distingue entre lo que se entiende por “concepto” y “concepción”. Ellos emplean la palabra concepto en el sentido de Vergnaud (1993). Para él, un concepto matemático no puede reducirse a su definición, sino que hay que tener en cuenta las situaciones en que se pone en juego, concibe un concepto como una terna de tres conjuntos (S, I, L), donde S representa al conjunto de situaciones que dan sentido al concepto (la referencia), I al de invariantes sobre los que reposa la operacionalidad de los “esquemas”¹ (el significado) y L al de formas de lenguaje y de no lenguaje que permiten representar simbólicamente el concepto, sus propiedades, las situaciones y los procedimientos de tratamiento (el significante).

Estudiar el desarrollo y el funcionamiento de un concepto, al curso del aprendizaje o en el momento de su utilización, es necesariamente considerar esos tres planos a la vez. No hay en general una biyección entre significantes y significados, ni entre invariantes y situaciones. No se puede reducir el significado ni a los significantes ni a las situaciones. (Vergnaud, 1993, p. 96)

Por otra parte, se emplea el término concepción en el sentido de Artigue (1990), es decir, en un sentido local. Es importante resaltar como lo anota Vallecillos (1996), que

[...] en la investigación didáctica sólo podemos tener acceso a unos aspectos concretos de los conocimientos efectivamente construidos por el sujeto, por tanto, para nosotros las concepciones de los estudiantes son la suma de los conocimientos y creencias puestas en juego y manifestadas en las situaciones concretas locales, de

1. Vergnaud (1993, pag. 89) define un esquema como “la organización invariante de la conducta por una clase de situaciones dada”. Además agrega que es en estos esquemas en donde hay que buscar los conocimientos-en-acto o teoremas en acto del sujeto, es decir los elementos cognoscitivos que permiten a la acción del sujeto ser operatoria”.

observación o evaluación que les han sido planteadas. (Vallecillos, 1996, p.119)

Las ideas que se plantearon en los párrafos anteriores nos dan una idea de base para plantear nuestro marco de análisis con respecto al tema del aprendizaje de los estudiantes. Entonces, para nosotros hablar de los conocimientos de los estudiantes es imponernos la tarea de describir tan detalladamente como podamos las ideas que creemos que tienen nuestros estudiantes acerca de los temas particulares que se quieren analizar. Por supuesto, estas descripciones se harán con base en la experiencia que se ha tenido, al trabajar durante varios semestres, con este tipo de estudiantes. Haremos dos tipos de descripciones: por un lado, descripciones del “qué tiene” el estudiante antes de poner en funcionamiento las situaciones que se van a aplicar, y por otro las descripciones del “qué necesita” para resolver las tareas que se les piensa proponer.

Por otra parte, vamos a considerar de manera informal la definición de concepto de Vergnaud, para aproximarnos al análisis de las situaciones que se van a aplicar. La idea es considerar una triada (Situaciones problemáticas (SP); Errores dificultades y obstáculos (EDO), y sistemas de representación (SR)) en el siguiente sentido:

- En primer lugar, tendremos un conjunto de situaciones problemáticas de referencia bajo las cuales vamos a comentar los conceptos relacionados con los tests de hipótesis, que se trabajan en cada situación. Para ello, vamos a hacer tan explícito como sea posible, los conceptos y los procedimientos que se ponen en juego y también las interrelaciones que se abordan entre estos.
- En segundo lugar, tendremos un conjunto descriptivo de errores, de dificultades y de formas de conocimiento que tienen los estudiantes con relación a cada situación específica que se les plantea. Somos conscientes de que esta no es precisamente la descripción del conjunto hipotético de “conocimientos-en-acto” (conjunto de invariantes de Vergnaud) de lo que el sujeto “sabe en la acción” (conocimientos que puede usar en la práctica pero que no los puede articular formalmente), sin embargo, la encontramos muy útil para caracterizar la situación problemática y para adquirir un conocimiento más profundo de los problemas de aprendizaje que tienen los estudiantes.
- Finalmente y en tercer lugar, queremos considerar los sistemas de representación externos como una manera de aproximarnos a la tercera dimensión que plantea Vergnaud, es decir, a la de las formas de lenguaje y de no lenguaje que permiten representar simbólicamente el concepto, sus propiedades, las situaciones y

los procedimientos de tratamiento. En este caso haremos uso de la idea planteada en Janvier (1987).

En lo que resta de esta parte, vamos exponer con más detalles los elementos que se esbozaron en los dos últimos puntos: errores, dificultades y obstáculos y sistemas de representación.

ERRORES Y DIFICULTADES

Dada la complejidad de la comprensión en matemáticas no es de extrañar que la construcción de conocimientos estadísticos que realiza un sujeto no coincida, en general, con la de otro sujeto. En realidad, el hecho observable de que un alumno se equivoque cuando se le pide realizar ciertas tareas, apoya la afirmación de que no todos los sujetos construyen los mismos conocimientos matemáticos sobre un mismo objeto matemático. Como consecuencia, la educación matemática ha dirigido en el terreno de los errores, dificultades y obstáculos, gran parte de la investigación teórica y experimental. En particular, la investigación en educación estocástica también puede utilizar la estrategia de análisis de errores. Como dice Radatz (1980), esta es “una estrategia prometedora para aclarar cuestiones fundamentales del aprendizaje ” (p. 16). También Rico (1995b) destaca la importancia del análisis de los errores en la educación matemática y Borassi (1987, p. 7) presenta esta estrategia “como un recurso motivacional y como punto de partida para la exploración matemática creativa, implicando valiosas actividades de planteamiento y resolución de problemas”.

Cuando un alumno no proporciona respuestas adecuadas con respecto a lo que normativamente se considera correcto (exceptuando los casos en que pueda decirse que se trata de una simple distracción), se dice que le resulta “difícil” la tarea en cuestión. Centeno (1988, p. 144) define una dificultad como “algo que impide ejecutar bien o entender pronto una cosa. Las dificultades pueden proceder de diversas causas, relacionadas con el concepto que se aprende, con el método que utiliza el maestro, con la preparación anterior del alumno o con su propia disposición para aprender.”

Obstáculos

Una de las problemáticas que se plantea en la educación matemática es que algunas de las concepciones que permiten resolver un conjunto de tareas en términos adecuados, son limitadas o inapropiadas cuando se aplican en casos más generales, casos donde el sujeto muestra una resistencia al cambio, a sustituir su forma de conocimiento. En estas circunstancias se habla de la existencia de un obstáculo que puede explicar la existencia de errores y dificultades especiales. Brousseau (1983) describe las siguientes características de los obstáculos

- un obstáculo es un conocimiento, no una falta de conocimiento.
- el sujeto utiliza este conocimiento para producir respuestas adaptadas a un cierto contexto que encuentra con frecuencia.
- cuando se usa este conocimiento fuera de este contexto se producen respuestas incorrectas; una respuesta universal exigirá un punto de vista diferente.
- el sujeto se resiste a las contradicciones que el obstáculo le produce y al establecimiento de un conocimiento mejor.
- después de haber notado su inexactitud, las respuestas incorrectas se siguen manifestando de manera esporádica.

Brousseau identifica tres tipos de obstáculos:

Obstáculos ontogénicos. Son propios de las características del desarrollo mental del estudiante, es decir, debidas a las limitaciones impuestas por la capacidad cognitiva del sujeto en el momento del aprendizaje. Por ejemplo, algunas veces los profesores suponen que los estudiantes poseen la capacidad de utilizar un lenguaje simbólico específico, tales como notaciones utilizadas en la estadística. Sin embargo, es posible que algunos errores detectados que tienen relación con la notación empleada, puedan tener su origen en un desarrollo deficiente de la capacidad de uso del lenguaje formal necesario.

Obstáculos didácticos. Son consecuencia de elecciones didácticas hechas para establecer una situación de enseñanza. Por ejemplo, la introducción de nuevo simbolismo para representar un concepto o un procedimiento, puede originar errores debido a que posiblemente sea necesario que los estudiantes trabajen previamente con ejemplos concretos.

Obstáculos epistemológicos. Están relacionados intrínsecamente al propio concepto y contienen parte del significado del concepto. En el caso de la estocástica, un ejemplo que se menciona con frecuencia es el de las circularidades que se presentan en las diferentes definiciones del significado de probabilidad (clásica, frecuentista y subjetiva) que dieron lugar en un momento dado de la historia, a una formulación axiomática con la cual se resolvía este conflicto.

Algunas formas de pensamiento prevalentes en la estocástica

Muchos de los trabajos de investigación que se han realizado en esta área se han concentrado en estudiar cómo se comprenden o mal comprenden ideas particulares de la estadística y la probabilidad. Algunos ejemplos de estas

formas de pensamiento prevalentes que son inconsistentes comparadas con un aprendizaje establecido como normativamente correcto, son presentadas en Shaughnessy (1991) y también en Scholz (1991). Mencionamos las siguientes:

Representatividad. La gente estima probabilidades de eventos con base en qué tan bien un resultado representa algún aspecto de la población de referencia. Concretamente la gente piensa que una muestra o debe reflejar la distribución de la población de referencia o debe reflejar el proceso por el cual los eventos aleatorios son generados. Por ejemplo, bajo esta heurística, una secuencia aleatoria de seis nacimientos en un hospital del tipo HMMHMM es más probable que una secuencia HHHHHM (H: hombre y M: mujer). Otro ejemplo dice que obtener un 70% de nacimientos de hombre es igualmente probable en una muestra de 10 nacimientos que en una muestra de 1000 nacimientos.

Falacia del jugador. La heurística de la representatividad también permite ver el azar como un proceso autocorrectivo. Por ejemplo, después de una gran racha de caras se piensa que es más probable que salga un sello a que vuelva a salir una cara.

Falacia de la proporción base. La gente ignora los tamaños relativos de subgrupos de poblaciones cuando juzga la probabilidad de eventos contingentes involucrando los subgrupos. Por ejemplo, cuando se juzga la probabilidad de que un estudiante sea ingeniero, la gente ignora la proporción de base y en vez de ello apoya su argumento con base en la personalidad del estudiante.

Disponibilidad. Cuando la gente estima la probabilidad de un evento con base en qué tan fácil es para ellos traer a la mente una instancia particular del mismo, está empleando esta heurística. Es muy corriente encontrar este tipo de razonamiento en la vida cotidiana. Por ejemplo, si usted pocas veces ha conocido que en determinado sector de la ciudad hay problemas de inseguridad, usted seguramente piensa que ese sector de la ciudad es seguro.

Falacia de la conjunción. La conjunción de dos eventos correlacionados se juzga que es más probable que la de cualquiera de los dos de manera independiente. Por ejemplo, se juzga que es más probable el evento tener más de 55 años y sufrir un paro cardíaco que simplemente sufrir un paro cardíaco.

Aproximación al resultado. La gente utiliza un modelo de probabilidad para tomar decisiones acerca de un simple evento en vez de considerarlo como haciendo parte de una serie de eventos. Por ejemplo, eventos con pro-

habilidades mayores de 70% de ocurrencia se consideran como eventos seguros, mientras que eventos con probabilidades de 30% se consideran como eventos imposibles.

SISTEMAS DE REPRESENTACIÓN

La premisa fundamental sobre la cual parte Kaput (1987) para plantear la necesidad de proponer una teoría de los sistemas de representación en matemáticas es que

los fenómenos básicos de aprendizaje y aplicación de las matemáticas son tratados con representación y simbolización porque estos son simultáneamente el corazón del contenido matemático y de las cogniciones asociadas con la actividad matemática. (p. 22)

Kaput reconoce que la parte más dura del desarrollo de esta teoría está en establecer la realidad psicológica/lingüística de los constructos asociados con la creación de símbolos externos y su coherencia con las estructuras y procesos cognitivos internos. Por ello recomienda primero, de manera informal y tal vez a manera de justificación, realizar un examen del contenido matemático en sí mismo desde una perspectiva representacional sin mirar las dimensiones curriculares, lingüísticas o psicológicas. De la propia revisión que hace Kaput (1987, pp. 23-25) concluye que “la mayoría de resultados vistos como fundamentales en matemáticas son fácilmente clasificables como representacionales.”

La mente humana ha mostrado ser considerablemente efectiva para manejar ideas extremadamente complejas, tanto a nivel abstracto como concreto. Esta potencia, se puede atribuir en parte, a la interacción entre dos fuentes de la organización de la experiencia: por un lado a nuestra memoria de largo término y por otro, a nuestra habilidad para explotar los medios físicos de organización de la experiencia. En el caso de la experiencia estadística y matemática, se utilizan *sistemas de representación* o formas inherentes de exteriorizar estructuras conceptuales, en las cuales se reflejan incluso, nuestras propias creencias personales.

En Janvier (1987), Kaput sostiene que cualquier concepto de representación debe involucrar dos entidades relacionadas pero funcionalmente separadas: el mundo representante y el mundo representado. De aquí concluye que cualquier especificación particular de una representación debe describir los siguientes cinco entidades: (1) el mundo representado, (2) el mundo representante, (3) los aspectos del mundo representado que están siendo representados, (4) los aspectos del mundo representante que están haciendo de representantes y (5) la correspondencia entre los dos mundos.

Para darle sentido a las argumentaciones que se puedan expresar a nivel cognitivo, es necesario referirnos a dos mundos diferentes: el mundo de las

operaciones mentales, el cual no es observable y por ello siempre es hipotético y el mundo de las operaciones físicas, que en general, es observable. Así pues, se pueden establecer dos tipos de interacción: una, del mundo de las operaciones mentales al mundo de las operaciones físicas, donde se proyectan las estructuras mentales existentes sobre algún material físico real o donde también se puede actuar sobre los ‘materiales físicos’ existentes para producir nuevas estructuras en forma de materiales (proceso de “escritura”); la otra, del mundo del físico al de la mente, donde se lee deliberadamente o se evoca pasivamente acciones reales que se dan en el mundo físico (proceso de “lectura”). De esta manera se habla entonces, de un proceso cíclico de “lectura” y “escritura” en el cual el investigador puede a partir de la observación de las acciones y operaciones físicas, lanzar conjeturas o hipótesis acerca de lo que sucede a nivel de las operaciones mentales. En la figura siguiente se ilustra este proceso.

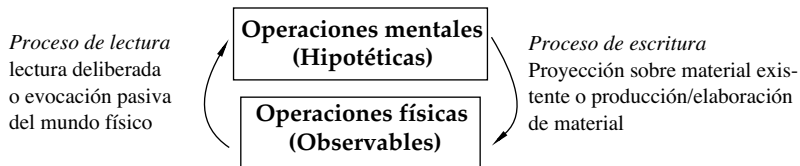


Figura 2.

La flecha orientada hacia abajo sugiere el proceso mediante el cual se proyecta un conocimiento cognitivo que se representa externamente para propósitos de comunicación. Por otra parte, la flecha orientada hacia arriba sugiere el proceso mediante el cual con base en “materiales físicos”, apoyamos las operaciones mentales.

5. ANÁLISIS DEL APRENDIZAJE DE LOS TESTS DE HIPÓTESIS

INTRODUCCIÓN

En esta parte consideramos el tema concreto del aprendizaje de los tests de hipótesis. La presentación se apoya en varios de los elementos conceptuales que se han expuesto anteriormente. En primer lugar, queremos hacer una caracterización de lo que podría ser el “estado de comprensión” de un estudiante promedio que entra a ver el curso “Estadística 2”. Luego comentamos, sin todavía hacer referencia específica a las situaciones problemáticas que se les van a plantear a los estudiantes, algunos de los tipos de errores que se manifiestan en los estudiantes y las posibles dificultades que los pueden originar.

COMPRENSIÓN ESTOCÁSTICA

Hablar de la comprensión estocástica que puede tener un grupo de estudiantes es muy difícil. Las descripciones que se van a hacer sólo pretenden ser descripciones hipotéticas de posibles estados de comprensión que se cree que tienen los estudiantes. Estas descripciones, como fue planteado en el marco conceptual, de la sección anterior, solamente pueden basarse en las actuaciones de los estudiantes y en la manera como manejan y expresan su conocimiento a través del lenguaje y la utilización de representaciones externas.

Ya que somos conscientes de las limitaciones y sobre todo de la complejidad de lo que significa describir un estado de comprensión, el esquema que se plantea a continuación y lo que podemos decir acerca de estos estados, sólo pretende ser una aproximación muy superficial con respecto a lo que idealmente se debería decir cuando hablamos de los estados de comprensión (a la manera de Sierpiska), con respecto a un concepto específico. El esquema de la descripción que se pretende realizar es el siguiente:

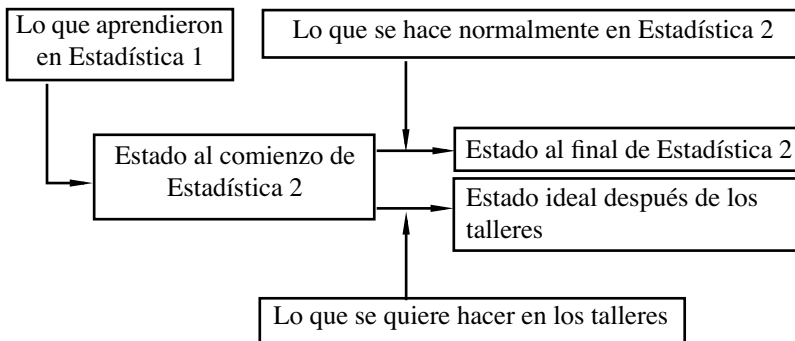


Figura 3.

Para las descripciones nos basamos por un lado, en una muestra de las pruebas que diseñaron los profesores que dictaron los cursos de Estadística 1 y en las correspondientes soluciones escritas de los estudiantes que dieron como respuesta a estas pruebas. Por otro lado, también nos basamos en la experiencia de los procesos de interacción que se han vivido con los estudiantes en el salón de clase. Cuando lo hemos considerado pertinente, se ha hecho referencia a algunos investigadores que describen con propiedad características de los estados que queremos describir.

Haremos una descripción del estado de comprensión dividiéndola en siete temas o nociones generales. Para cada noción, primero se hace una descripción breve que luego se complementa atendiendo al “qué tienen” y al “qué necesitan” los estudiantes para desarrollar los temas que se han propuesto trabajar en el diseño de las situaciones problemáticas. Estas nociones también se pueden identificar con algunas de las cinco unidades de análisis propuestas en el análisis de contenido. Son las siguientes:

- Nociones de población y muestra(datos)
- Medición, escalas y variables (datos)
- Lógica del contraste (formulación)
- Medidas de localización, de variabilidad y estadísticos de prueba (herramientas)
- Azar y probabilidad (distribuciones)
- Noción de distribución (distribuciones)
- Nivel de significación (análisis)

Estado de comprensión al comenzar Estadística 2

Nociones de población y de muestra

Estas dos nociones aparentemente son claras para la mayoría de estudiantes. Sin embargo, los exámenes parciales que usualmente se realizan en los primeros días de clase, muestran que hay cuestiones relacionadas con los procesos de obtención de muestras que reflejan que no hay una clara conceptualización. Por ejemplo, casos como el de la heurística de la representatividad o la falacia del jugador, mencionados en la parte anterior, suelen ocurrir en algunos de los estudiantes observados.

También es usual observar la dificultad para establecer una distinción entre lo que se considera población de estudio y lo que se considera población de datos, distinción que también se aplica al caso de muestras.

Con respecto al “qué tienen” podemos decir lo siguiente:

- Identifican que una muestra es un subconjunto de la población, pero no necesariamente discriminan entre una población (muestra) de datos y una población (muestra) de estudio.
- Reconocen que existen muestras representativas y muestras aleatorias. Sin embargo, algunos estudiantes piensan que toda muestra aleatoria es representativa o viceversa.
- Identifican algunos tipos de muestreos aleatorios tales como el muestreo aleatorio simple o el aleatorio estratificado y algunos incluso han trabajado con tablas de números aleatorios.

Con respecto al “qué necesitan” podemos decir que deberían:

- Identificar las diferencias entre poblaciones (muestras) de datos y población (muestra) de estudio.
- Identificar que muchos de los métodos estadísticos de inferencia se basan en el supuesto de que las muestras de datos deben haber sido obtenidas de manera aleatoria.
- Distinguir claramente la diferencia entre aleatoriedad y representatividad.

Medición, escalas y variables

El común de estudiantes puede identificar claramente dos tipos de categorías de clasificación que son las más simples: la cuantitativa y la cualitativa. Suelen identificar que la categoría cuantitativa se asocia claramente con expresiones numéricas mientras que la cualitativa se asocia con expresiones literales.

Con respecto a las denominadas variables categóricas, la tendencia de los estudiantes es considerarlas como exclusivamente de tipo cualitativo,

aunque ellos también hayan trabajado con variables continuas que han sido categorizadas.

En cuanto al manejo de escalas y al concepto de estandarización, el estudiante ha tenido contacto con su manejo. Sin embargo, los documentos que se tienen de soluciones a pruebas escritas algunas veces reflejan que en el caso de la estandarización parece evidenciarse un conocimiento procedimental muy desligado de una apropiada comprensión conceptual de lo que es una puntuación estándar. Los estudiantes pueden estandarizar o desestandarizar un dato pero no entienden que significado tiene exactamente el resultado que obtienen.

Con respecto al “qué tienen” podemos decir lo siguiente:

- Identifican y saben construir gráficos para representar variables continuas (histogramas) y discretas (gráficos de puntos o de barras).
- Saben construir tablas de distribuciones de frecuencias para variables cuantitativas y cualitativas y en general, representar estas distribuciones en gráficos (histogramas, “tortas” y diagramas de bloques). Sin embargo, algunas veces no se establecen diferencias entre los tipos de gráficos que son más apropiados para las variables consideradas.
- Se ha trabajado más con variables continuas que con otro tipo de variables.

Con respecto al “qué necesitan” podemos decir que deberían:

- Mejorar los procesos de identificación y discriminación de representaciones gráfica, simbólicas, numéricas y verbales de los conceptos de variable y escala. En particular, enfatizar en el manejo de variables categóricas y de conteo.
- Que trabajen con contextos de la vida cotidiana donde se enfoquen el manejo e interpretación de la medición, la escala y el uso de variables.

Lógica del contraste

La formulación de las hipótesis estadísticas es uno de los aspectos donde los estudiantes tienen más dificultades. Esto no debe sorprendernos si recordamos que detrás de la lógica que subyace el manejo y la aplicación de un contraste de hipótesis hay un pasado histórico y epistemológico en el que se encuentran, por ejemplo, las discusiones entre la posición bayesiana y la de Fisher, e incluso entre la Neyman-Pearson y Fisher.

Pocas veces el estudiante es suficientemente consciente, cuando está aplicando un contraste, de que la distribución bajo la cual está trabajando es la que se supone bajo la hipótesis nula. Ante las dificultades que implica la

formulación de la hipótesis nula contra una alternativa el estudiante prefiere la memorización de una regla que en pocas palabras puede expresarse así: *la hipótesis nula siempre debe llevar un “=” un “ \leq ” o un “ \geq ” y la alternativa un “ \neq ” o un “ $<$ ” o un “ $>$ ”*. Este tipo de reglas incluso es recomendada en algunas ocasiones por el profesor y por los textos. Con ello se favorece una comprensión esencialmente procedimental de la lógica del contraste de hipótesis, que incluso niega la posibilidad de proponer hipótesis simples en la alternativa como se puede dar usualmente en la lógica que apoya la posición de Neyman-Pearson.

Con respecto al “qué tienen” podemos decir lo siguiente:

- Aunque con algunas dificultades, en general, identifican y discriminan formulaciones de hipótesis estadísticas que vienen dadas en alguna de las siguientes formas:
 - a) H_0 : Parámetro $\geq c$ contra H_a : Parámetro $< c$
 - b) H_0 : Parámetro $\leq c$ contra H_a : Parámetro $> c$
 - c) H_0 : Parámetro $= c$ contra H_a : Parámetro $\neq c$, donde c representa cualquier número real.
- No tienen suficiente claridad para identificar hipótesis unidireccionales a partir de las expresiones verbales de los problemas. Los casos en donde la hipótesis de investigación está explícitamente formulada en los problemas en términos de ‘mayor que’ o ‘menor que’ los pueden identificar más fácilmente.
- Hay evidentemente una gran dificultad para argumentar, en términos de la lógica del test de hipótesis, porqué en la hipótesis nula no puede ocurrir el caso “parámetro $\neq c$ ”. Sus argumentaciones suelen reflejar formas de comprensión de la lógica del contraste en términos puramente procedimentales.
- Con respecto a los parámetros con los que se ha tratado, sólo se ha trabajado con la media aritmética y con la diferencia de medias aritméticas. Aunque También han identificado la desviación estándar o la proporción como parámetros no han trabajado procesos de inferencia relacionados con estos parámetros.

Con respecto al “qué necesitan” podemos decir que deberían:

- Identificar y discriminar argumentos de tipo conceptual y procedimental que expliquen la lógica de la aplicación de la regla: *la hipótesis nula siempre debe llevar un “=” un “ \leq ” o un “ \geq ” y la alternativa un “ \neq ” o un “ $<$ ” o un “ $>$ ”*.
- Distinguir entre lo que es un parámetro, lo que es la variable que origina un parámetro y lo que es el estimador mismo de un parámetro.

- Identificar y discriminar situaciones donde pueden ocurrir los errores de tipo I y II. Por ejemplo, con problemas donde se proponga el manejo de hipótesis alternativas simples donde se pueden calcular e interpretar concretamente errores de tipo II.
- Poder generalizar la lógica de los tests de hipótesis a los nuevos tipos de tests que se les presenten.

Azar y probabilidad

La noción de azar y algunos de los conceptos de probabilidad hacen parte de las bases de comprensión que poseen los estudiantes que inician el curso de Estadística 2. Suponemos, por un lado, que el tipo de experiencias que han vivido los estudiantes con estas nociones, como son algunos juegos de azar y algunas experiencias didácticas que conceptualmente podemos considerar como experimentos de simulación, no parecen ser suficientes para generar estados de comprensión sobre el azar y la probabilidad que se puedan considerar como normativamente correctos.

Un aspecto muy interesante que quisimos observar de manera informal en los estudiantes, es la confirmación de algunos de los resultados que se han publicado en la literatura de investigación estocástica sobre este asunto. Por ejemplo, en Shaughnessy (1993), se referencian varios tipos de pensamientos prevalentes, de concepciones erradas o de falacias alrededor de estos conceptos. Cuestiones tales como las falacias del jugador y de la conjunción y las heurísticas de la disponibilidad y de la aproximación al resultado, se han encontrado que efectivamente se dan con cierta frecuencia en razonamientos de los estudiantes.

Con respecto al “qué tienen” podemos decir lo siguiente:

- Saben aplicar las definiciones clásica y estadística de la probabilidad a eventos asociados con experimentos aleatorios del tipo: “lanzar una, dos, tres... veces una moneda”, “lanzar uno, dos, tres... dados”, etc.
- Saben calcular las probabilidades de complementos y de uniones de eventos relacionados con experimentos aleatorios similares a los mencionados en el punto anterior.
- Recurren con poca frecuencia a la utilización de árboles de probabilidad para construir los espacios muestrales asociados con los tipos de experimentos aleatorios relacionados antes.
- El concepto de independencia de eventos no lo identifican con el tema de probabilidad condicional o con la relación entre el producto de probabilidades marginales y conjuntas de eventos.

Con respecto al “qué necesitan” podemos decir que deberían:

- Identificar y relacionar el concepto de independencia de eventos con las definiciones de probabilidad condicional, probabilidad conjunta y producto de probabilidades marginales, en particular en el contexto de problemas donde se manejen tablas de doble entrada (p.e. tablas de contingencia).
- Poner en juego problemas de cálculos de probabilidades de experimentos aleatorios que los lleven a identificar y discriminar contextos diferentes a los de monedas y dados.
- Identificar características típicas y características que no son propias de los experimentos de tipo binomial y multinomial.

Nivel de significación

Este es uno de los conceptos del razonamiento estocástico cuya comprensión ha sido muy discutida. Por ejemplo, uno de los aspectos de comprensión sobre el cual se han llevado a cabo múltiples investigaciones (Oakes, 1986; Pollard y Richardson, 1987 y Falk y Greenbaum, 1995, entre otros) tiene que ver con la interpretación de lo que es un resultado significativo. Entre otros, Falk (1986) señala la confusión que se encuentra corrientemente entre estudiantes e investigadores en la definición de lo que es el nivel de significación. Concretamente el nivel de significación definido como $\alpha = P(\text{rechazar } H_0 \text{ dado que } H_0 \text{ es cierta})$ es usualmente interpretado como $P(H_0 \text{ cierta dado que se ha rechazado } H_0)$. Falk atribuye como una de las posibles causas de este error, el lenguaje utilizado en la definición del nivel de significación, es decir, como la “probabilidad de error de tipo I”. Explica que como en esta expresión no se indica explícitamente que se está trabajando con una probabilidad condicional, el estudiante supone que es posible definir un suceso condicional.

Vallecillos (1996), distingue cuatro aspectos diferenciados de la comprensión de este concepto que intervienen en los contrastes de hipótesis:

- El contraste de hipótesis como un problema de decisión. Al respecto algunos estudiantes interpretan los errores de tipo I y II como sucesos complementarios, por consiguiente la probabilidad de cometer alguno de los errores sería 1.
- Nivel de significación como riesgo del decisor. Algunos alumnos creen que el cambio en el nivel de significación no afecta al riesgo de error en la decisión.
- Interpretación de un resultado significativo. Un resultado estadísticamente significativo no implica que necesariamente sea relevante desde el punto de vista práctico, algunos estudiantes confunden ambas cosas. En otras ocasiones piensan que un resultado significativo es el que corrobora la hipótesis nula.

- Las probabilidades de error y la relación entre las mismas. Probablemente son muy pocos los estudiantes que tienen y tendrán una comprensión de α y β como probabilidades condicionales y menos de la dependencia de β como función del parámetro desconocido y de las relaciones entre α y β . Por ejemplo, muchos de los estudiantes suprimen la condición en la probabilidad condicionada que se emplea para definir α .

En particular, con respecto al “qué tienen” los estudiantes que entran al curso de estadística 2, podemos decir lo siguiente:

- Han trabajado muy poco con la definición condicional de los errores de tipo I y II. Estas definiciones tienden a ser memorizadas de manera verbal y a no ser relacionadas con la lógica del contraste de hipótesis.
- No han trabajado con el concepto de p-valor.
- Los problemas de tests de hipótesis que han trabajado no propician la reflexión en torno de la relación entre tipos de errores y nivel de significación.
- Han trabajado con intervalos de confianza para la media de una población. Al respecto, por ejemplo, la representación simbólica para el nivel de confianza de un intervalo dada por $(1 - \alpha)\%$ les genera conexiones con el nivel de significación α , que los lleva a hablar de tests de hipótesis “con niveles de seguridad” de $(1 - \alpha)\%$.

Con respecto al “qué necesitan” podemos decir que deberían:

- Identificar y discriminar más claramente las diferencias y similitudes entre los conceptos de nivel de significación, p-valor, tipos de error y nivel de confianza.
- Propiciar situaciones que pongan en juego cada uno de los cuatro aspectos de la comprensión relacionados con el nivel de significación mencionados anteriormente por Vallecillos.

Medidas de localización, de variabilidad y estadísticos de prueba

Con respecto a las medidas de tendencia central, la mayoría de estudiantes reconocen y diferencian la media aritmética, la mediana y la moda; no identifican otras medidas de tendencia central tal como la media armónica o la media geométrica que usualmente son menos conocidas. En el manejo de la media aritmética se evidencian dificultades para calcularla de manera agrupada, por ello suponemos que no hay una clara conceptualización de lo que es un promedio. Por ejemplo, a veces confunden la frecuencia de un

dato con el dato mismo o dividen la suma de los datos por el número de categorías y no de los datos de la muestra.

En cuanto a las medidas de posición, son bien reconocidos, los valores máximo y mínimo de una muestra, pero con menos frecuencia los percentiles y en particular los cuartiles.

A nivel procedimental se evidencia en algunos estudiantes la falta de destreza para calcular la desviación estándar, e incluso no saben expresar correctamente su expresión simbólica. Aparentemente los estudiantes reconocen la importancia de la noción de dispersión como relevante para la descripción de una distribución de datos. Sin embargo, cuando se les pide comparar algunas distribuciones varios de ellos no caen en la cuenta de que la información que proporcionan las medidas de tendencia central, da una descripción incompleta. Entonces con respecto a esta noción nos enfrentamos con estudiantes que además de no tener una comprensión conceptual que al menos de cuenta de diferencias entre distribuciones con medidas de tendencia central similares, tienen dificultades de orden procedimental para calcular una desviación.

Con respecto al “qué tienen” podemos decir lo siguiente:

- Saben, en general, calcular un promedio muestral, mientras no se pida que lo hagan con datos que no están agrupados. La evidencia de que para muchos no es claro como calcular un promedio de manera agrupada, sugiere que incluso un concepto tan sencillo como el de promedio, no está suficientemente conceptualizado.
- Con respecto a la desviación estándar su comprensión es más deficiente que la de la media aritmética. Por ejemplo, a nivel procedimental se encuentran casos donde los estudiantes no identifican correctamente la expresión simbólica de las desviación estándar y no discriminan entre la desviación y la varianza. Muy pocos estudiantes pueden dar explicaciones de porque existen dos varianzas o desviaciones típicas, una donde se divide por n y la otra donde se divide por $n-1$. No saben que significa que un estimador sea insesgado. Simplemente piensan que la desviación donde se divide por n es poblacional y que la otra (en la que se divide por $n-1$) es la muestral.
- En contextos de inferencia acerca de medias o de diferencia de medias se ha trabajado para que se distinga la diferencia entre lo que es el estimador de un parámetro y el parámetro mismo, sin embargo, no es raro encontrar casos donde se confunden a nivel de la representación simbólica.

- Dentro de los tests de hipótesis también se ha trabajado con los cocientes

$$\frac{\bar{x} - \mu}{\sqrt{\frac{S^2}{n}}} \qquad \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

pero su distribución no se identifica con otros modelos diferentes del normal. La conexión con el modelo t-student se establece en el curso de estadística 2.

Con respecto al “qué necesitan” deberían:

- Identificar y discriminar entre varianza insesgada y sesgada, y entre varianza poblacional y muestral.
- Identificar y utilizar la noción de dispersión en el manejo de los test de hipótesis. Por ejemplo, proponiéndoles problemas de tests de hipótesis donde tengan que manejar reglas de decisión que no se dan de manera estandarizada.
- Discriminar y generalizar el concepto de promedio aritmético simple, ponderado y agrupado. Por ejemplo, proponiéndoles problemas donde se deban calcular e interpretar promedios en forma agrupada.

Noción de distribución

Si nos referimos simplemente a una tabla de frecuencias puede encontrarse que confunden la variable con los valores. Por ejemplo, en una tabla de frecuencias promedian los valores o las frecuencias pero no identifican que ambas columnas de la tabla son necesarias para realizar cálculos.

Por otra parte, uno de los problemas que más se evidencia, se refiere a las dificultades que tienen los estudiantes para distinguir entre lo que es una distribución poblacional, una distribución de una muestra cualquiera y la distribución muestral de un estadístico tal como la media aritmética, para realizar un test de hipótesis. Al respecto, en Schuyten (1991), encontramos algunas explicaciones que justifican estas dificultades. Entre ellas se puede mencionar la que se refiere a los diversos niveles de concreción en un concepto en inferencia: mientras en la estadística descriptiva la unidad de análisis es una observación (un sujeto o un objeto) y calculamos la media aritmética de tales objetos, en inferencia, estamos interesados, entre otros problemas, por obtener información sobre la esperanza matemática de la media aritmética. Lo anterior significa que una muestra sobre la cual se puede calcular un promedio, se convierte en una observación de otra población diferente (la de todas la medias posibles de un tamaño dado). Esto introduce una gran dificultad conceptual que es necesario tener en cuenta.

Con respecto al “qué tienen” podemos decir lo siguiente:

- Usualmente reconocen que el promedio muestral, sigue un modelo de distribución normal, sin embargo, no identifican claramente esto como resultado de que la distribución muestral del promedio tiende a tener la forma normal cuando n es grande o como resultado de aplicar el teorema del límite central.
- Se ha trabajado con la tabla de la distribución normal. Saben buscar puntuaciones ‘ z ’ correspondientes a valores de α dados tales como, 10%, 5% o 1%. el sentido inverso. Hay dificultades para entender que se puede utilizar en el sentido inverso.
- En general distinguen las zonas de rechazo y de no rechazo de la hipótesis nula. Sin embargo, este manejo lo saben expresar mejor en un nivel de representación gráfico que en un nivel simbólico.
- El concepto de distribución muestral se ha trabajado acudiendo a ejemplos particulares donde se construyen distribuciones muestrales para la media o para la diferencia de medias. Sin embargo, como ya mencionó antes, este es uno de los conceptos más difíciles de asimilar. No parece que sea suficiente mostrar a través de ejemplos concretos, cómo se construye una distribución muestral. Un proceso de construcción de una distribución muestral lo pueden recordar a nivel procedimental, pero pocas veces logran identificarlo de manera concreta con el teorema del límite central.

Con respecto al “qué necesitan” podemos decir que deberían:

- Identificar y discriminar aspectos relacionados con las representaciones tabulares de las distribuciones, tales dominio y rango de las distribuciones.
- Identificar y diferenciar entre las diferentes formas de representación de las distribuciones. Por ejemplo, deben mirar las regiones de rechazo y de no rechazo en formas estandarizadas y no estandarizadas y en formas gráfica y simbólica.

Errores más frecuentes y posibles dificultades

I. Relacionados con el planteamiento y la formulación de hipótesis estadísticas

1. Hay casos donde los parámetros son confundidos con la variable dependiente o explicativa del problema. Posible dificultad:

- El estudiante piensa que un promedio sólo tiene sentido, cuando se puede calcular, y esto sólo se puede hacer sobre una muestra,

por esta razón el estudiante piensa que lo que se estudia como parámetro es la variable. En otras palabras el estudiante ve en el promedio muestral el indicador del comportamiento de la variable y entonces piensa que ésta es el parámetro de estudio. Generalmente un estudiante que comete este error no sabe responder claramente qué es un parámetro.

2. Se encuentran planteamientos de hipótesis donde se confunden los parámetros con los estimadores. Posibles dificultades:

- El estudiante piensa que no hay necesidad de hacer la distinción ya que el parámetro, al ser un valor desconocido que no se puede calcular, no hay que mencionarlo.
- También puede ser que el estudiante no tiene claro el sentido predictivo de los tests de hipótesis, es decir, no ha entendido que el objeto de análisis del problema es la estimación acerca de parámetros.
- También puede ser que no está tan familiarizado como debiera con la notación.

3. Situaciones donde no se identifica cuáles son las variables que intervienen en el problema. Posibles causas:

- Esto sucede cuando se presentan enunciados complicados con los que están poco familiarizados. La presentación de problemas con los que no están acostumbrados a trabajar, les genera un sentimiento de bloqueo.
- Es muy poco frecuente que en los problemas sobre inferencia que se les suele presentar se considere la pregunta de identificar la variable.

4. Hay tests unilaterales donde no reconocen cuál debería ser el sentido (izquierdo o derecho) del test. Posibles causa:

- Posiblemente porque no tiene ningún sentido para los estudiantes la hipótesis de investigación.
- Porque se piensa que la lateralidad se determina a partir de la hipótesis nula.
- Porque se piensa que la región de rechazo es la de aceptación y viceversa.
- Aplican sin considerar excepciones, la regla de que la hipótesis alternativa debe siempre coincidir con la hipótesis de investigación.

5. No se plantean las hipótesis estadísticas pero se hace el proceso de prueba. Posible causa:

- Porque saben que también se puede resolver el problema con base en intervalos de confianza, y en esta metodología de inferencia no es necesario plantear las hipótesis nula y alternativa.

II. Relacionados con la identificación de la población y la muestra

1. Confundir la muestra con la población o viceversa. Posibles causas:

- Este error sucede con más frecuencia en problemas donde los enunciados presentan de manera explícita, los tamaños de la población. El estudiante piensa que como éste dato se da en el enunciado, no es necesario discriminarlo del dato muestral.
- También sucede con frecuencia en enunciados donde se presentan dos poblaciones de investigación, pero donde la muestra solo puede aplicarse a una de las dos poblaciones.

2. Confundir datos históricos que deben interpretarse como parámetros, con datos muestrales.

- Si los datos históricos, son datos que se obtuvieron de algún estudio, seguramente son datos muestrales. Por esta razón parece poco lógico que se comparen dos muestras.

3. No es clara la diferencia entre muestras independientes y muestras relacionadas.

- No se sabe qué son muestras relacionadas, sólo han trabajado con muestras independientes. Los ejemplos de problemas de muestras relacionadas son sólo del estilo de “pre-post”, cuando eventualmente aparecen problemas de muestras aparejadas, las consideran como muestras independientes.

III. Relacionados con cálculos y aplicación de las herramientas

1. Se presentan fallas al sacar promedios y varianzas que necesitan calcularse de forma agrupada.

- Se enfrentan con este tipo de problemas con poca frecuencia.
- Confunden los valores de las frecuencias con los datos de la variable misma. Los datos que realmente representan los valores de la variable son las “marcas de clase”, pero estos no lo son para el estudiante.
- Los textos utilizados en la enseñanza presentan muy pocos problemas de inferencia donde se presentan estos casos.

2. No diferencian entre la varianza y la desviación estándar.

- Aunque esto parece que en principio se puede deber a una simple distracción, también se puede pensar que a nivel de la representación simbólica la diferencia entre una s y una s elevada al cuadrado no obliga a que se explicita una diferencia.

3. Se presenta un intervalo en la forma $[a, b]$ con $a > b$.

- La notación utilizada para los intervalos de confianza no resalta la importancia que tiene la noción de orden numérico.
- Es un tipo de representación simbólica que no han utilizado con frecuencia.

4. Se cometen errores algebraicos tales como: $\langle \bar{X} \pm EE(\bar{X}) \rangle \cdot K$, donde K es un valor que se encuentra en una tabla de distribuciones.

- Creen que se puede multiplicar en cualquier orden sin darle sentido a la prioridad de las operaciones.

5. Se comenten muchos errores de cálculo aritmético en la expresión del cociente z que define la estadística de prueba.

- Por ejemplo, restan el promedio mayor menos el menor cuando lo han debido hacer al contrario si hubieran sido consistentes con el planteamiento de las hipótesis, en otras palabras, se utiliza una diferencia muestral entre promedios en sentido inverso al que se ha planteado en las hipótesis estadísticas. También influye el hecho de que el cálculo de la estadística para muestras independientes es más largo. En este caso, el valor del cociente (t o z) se identifica directamente con el valor del error estándar de la diferencia de promedios muestrales.

IV. Relacionados con el uso y aplicación de distribuciones

1. Se presentan fallas en el manejo de las tablas de la distribución normal.

- Por ejemplo, no se establece la diferencia entre el valor percentil y la probabilidad. Puede deberse a la diferencia que hay entre las formas de las representaciones tabulares junto con la tendencia del estudiante a memorizar el proceso sin detenerse a mirar las características de la representación.

2. No se reconoce la diferencia entre una distribución estándar y una que no es estándar.

- Por ejemplo, al representar gráficamente una distribución no establecen si se trata de una distribución estándar o no. Cuando

se pide que mejoren la representación pueden colocar un cero como media de la distribución a una distribución con una media diferente de cero. Lo anterior puede atribuirse a que se enfatice en el procedimiento de test de hipótesis siguiendo estrictamente un algoritmo donde se comparan solamente los valores estandarizados.

V. Relacionados con el análisis e interpretación de los resultados

1. No se interpreta de manera correcta el significado del error de tipo I.

- Este error está muy relacionado con la concepción que se tenga de lo que es el nivel de significación. Por ejemplo, el no tener en cuenta la condicionalidad de la probabilidad de este error, hace que el estudiante la suponga como una probabilidad que puede ocurrir a posteriori.

2. Se toma una decisión, pero no se interpreta el resultado en términos del enunciado del problema.

- En la ocurrencia de este error puede pensarse por un lado, que existe la necesidad de tener contextos que tengan mucho más interés para los estudiantes que los motive a trabajar en términos del mismo. Sin embargo, por otro lado, las formas como se presentan los enunciados deben enfatizar este aspecto.
- Otra razón que se puede atribuir puede encontrarse en el predominio de una actitud procedimental en el estudiante ante el problema planteado, algo así como que en últimas lo que interesa saber es si hay que rechazar o no la hipótesis nula y con eso se resuelve el problema.

3. A veces se toma una decisión y se interpreta, sin embargo la interpretación se hace al contrario.

- En este error parece muy probable que el estudiante piensa que un resultado significativo es el que precisamente favorece la hipótesis nula.

Reflexión final acerca del estado de comprensión

El conocimiento que tienen los estudiantes acerca de lo que es un test de hipótesis es muy deficiente. Se perciben fallas en varios sentidos, entre las cuáles se pueden mencionar que no es clara la lógica que subyace a los tests de hipótesis, que hay dificultades en el manejo de distribuciones e incluso que se perciben en algunos de los estudiantes, la falta de destrezas para realizar cálculos aritméticos elementales. Esto es de esperar, si por un lado consideramos que en general la mayoría de estudiantes sólo ha tenido con-

tacto con los temas de la estocástica durante un semestre, y por otro, si consideramos toda la complejidad de lo que significa comprender lo que es un test de hipótesis. Para nosotros, un test de hipótesis es un “megaconcepto” para el que difícilmente se pueden establecer suficientes conexiones entre las diferentes unidades conceptuales que lo componen, durante el lapso de tiempo que se tiene para ello.

Nuestra visión con respecto a los estudiantes que entran a estudiar el curso de Estadística 2, es que ellos entienden los tests de hipótesis simplemente como una forma de hacer inferencia acerca de una media poblacional. Establecen conexiones débiles entre cuestiones tales como lo que es un modelo de distribución normal, un estadístico de prueba, zonas de rechazo y aceptación, etc. Y aunque es posible que muchos tengan suficiente destrezas procedimentales para realizar tests de hipótesis acerca de medias o diferencia de medias para llegar a respuestas correctas de problemas propuestos, posiblemente no sepan interpretar correctamente el significado mismo de la respuesta obtenida a la luz de aspectos tales como nivel de significación, p-valor, error de tipo I, etc.

ANÁLISIS EN TÉRMINOS DE REPRESENTACIONES

Nuestra visión acerca de la comprensión en relación con el uso de sistemas de representación externos tiene una premisa básica: para un concepto dado, entre más ricas y variadas sean las traducciones que pueda realizar apropiadamente un estudiante entre los diferentes sistemas de representación externo, tanto mejor será la comprensión de dicho concepto.

Con base en el esquema que presenta Janvier (1987), y que fue detallado en el capítulo de análisis de contenido, presentamos a continuación el ejemplo de un análisis donde se ilustran algunas de las traducciones entre diferentes sistemas de representación externos que se ponen en juego alrededor de los conceptos de nivel de significación y p-valor. Hemos elegido referirnos a éstos conceptos, por considerar que éstos son de vital importancia en el análisis y la interpretación de los resultados de un test de hipótesis.

Traducción de tablas a situaciones (descripciones verbales). Este tipo de traducción usualmente ocurre de manera indirecta, cuando se trata del concepto nivel de significación. Generalmente dado el valor de un nivel de significación α , se realiza una lectura en una tabla de distribuciones para establecer un valor crítico (por ejemplo, para encontrar un valor $z(\alpha)$, en el caso de una distribución normal). Dicho valor posiblemente se utilice para tomar la decisión sobre un test y para posteriormente comentar si se rechazó o no, con el nivel de significación especificado. Por otra parte, cuando se trata de un p-valor, es muy común que se deba realizar la lectura de una tabla de resultados. Sin embargo, esta tabla de resultados segura-

mente no será la de una distribución, sino que posiblemente será la de una salida de computador, en la que se deberá interpretar cual es la significación del resultado. Este tipo de interpretación puede ocurrir con mucha frecuencia en la práctica y sin embargo, es una de las que más se le dificulta al estudiante.

Traducción de situaciones (descripciones verbales) a tablas. En este caso el procedimiento de “medir” mencionado por Janvier (1987) debe verse más bien como un procedimiento de “calcular y organizar información”. Aunque se pueden presentar tareas en las que el propósito sea construir tablas de distribuciones que de alguna manera tendrán implícita o explícitamente niveles de significación y p-valores. Aunque este tipo de tareas no es usual en el nivel de estudiantes que queremos estudiar, creemos que su realización podría contribuir a mejorar la comprensión de éstos conceptos. En realidad, la primera situación problemática considera este tipo de traducciones, a través del uso de la calculadora gráfica.

Traducción de gráficas a situaciones (descripciones verbales). Las gráficas muchas veces se utilizan como ‘un puente’ para interpretar los resultados de un test de hipótesis. Por ejemplo, el esbozo de una curva acampanada que señala las regiones de rechazo y de no rechazo de una hipótesis nula, en general, le facilita al estudiante la interpretación del resultado de un test. Además, en las gráficas se pueden representar no sólo el nivel de significación y el p-valor, sino también otros conceptos relacionados tales como los que tienen que ver con la formulación de las hipótesis.

Traducción de situaciones (descripciones verbales) a gráficas. Implican procedimientos de dibujar. Podría decirse que una comprensión adecuada de lo que es el nivel de significación o el p-valor, se refleja, como condición necesaria, en una adecuada representación gráfica de dichos conceptos. Por otra parte, en las representaciones gráficas en las que se manifiestan errores se pueden evidenciar algunas de las concepciones erradas de los estudiantes que son más típicas.

Traducción de situaciones (descripciones verbales) a fórmulas. Dado que nos estamos limitando a comentar lo que se refiere a nivel de significación o p-valor, el procedimiento “modelar” no tiene sentido en este tipo de Traducción, aunque si es muy corriente en estadística cuando se construyen modelos tales como los modelos de regresión. En todo caso consideremos por ejemplo, una expresión tal como: $\alpha = P(\text{rechazar } H_0 \mid H_0 \text{ es cierta})$ que no es una fórmula, sino más bien una representación simbólica. En ella encontramos de manera explícita el evento condicionante que indica el supuesto de que la hipótesis nula es cierta. Este punto trae a colación la concepción errada, referenciada en páginas anteriores, de considerar que el nivel de significación es $\alpha = P(H_0 \text{ cierta} \mid \text{se rechazó } H_0)$.

6. ELEMENTOS PARA EL ANÁLISIS DE LA ENSEÑANZA

INTRODUCCIÓN

En esta sección vamos a considerar algunos aspectos relacionados con la enseñanza. En primer lugar, se presentan algunas consideraciones acerca del análisis de los textos de enseñanza, después, acerca de las situaciones problemáticas que se presentan a los estudiantes y el papel del profesor en la enseñanza de la estadística y por último algunas cuestiones acerca del uso de la tecnología.

ANÁLISIS DE TEXTOS

La importancia de los textos como material didáctico de apoyo ha sido señalada por diversos autores. Por ejemplo, Romberg y Carpenter (1988, p. 868) afirman que “el libro de texto es visto como la autoridad del conocimiento y la guía del aprendizaje”. En el informe Cockcroft se menciona que

los libros de texto constituyen una ayuda inestimable para el profesor en el trabajo diario del aula; le ofrecen una estructura en cuyo seno cabe desarrollar el trabajo de la asignatura”. (Cockcroft 1985, p. 113)

Por otra parte, estamos de acuerdo con Goetz y Lecompte (1988) en que con el análisis de los libros de texto se pueden establecer diferencias entre los objetivos de un programa y las formas como éste se lleva a la práctica. En este sentido y también en términos de la utilidad pedagógica que representa la lectura de los textos de estadística para los estudiantes y para el profesor, justificamos el incluir dentro de la indagación metodológica propuesta en este libro, el análisis de los textos que se han utilizado en el curso de estadística que nos ocupa y también los del curso que le precede.

Dentro de la bibliografía disponible relacionada con la educación estocástica son muy escasas las referencias a trabajos de análisis de textos. En todo caso, encontramos en trabajos tales como los de Brewer (1986), Malara (1989) y más recientemente en Harwell (1994) cómo se pone de manifiesto el que los libros de texto siempre se han de emplear con cuidado, y en función de las necesidades de los alumnos. Por ejemplo, consideraciones tales como si vale la pena o no excluir un capítulo, cambiar el orden de la presentación de algunos temas específicos, introducir problemas especiales, etc. son de relevancia en la estructuración de un diseño

curricular. En últimas como dicen Ortiz et al. (1996, p. 20) al citar a Chevallard (1985), siempre deberemos ser conscientes de que “los libros de texto se convierten en la norma de progresión del conocimiento de los alumnos al ofrecer una concepción legitimada del saber a enseñar.”

También en Ortiz et al. (1996), encontramos una posición con relación al análisis de textos a la cual nos adherimos:

Los libros de texto son un medio cuyo análisis nos proporciona un conocimiento del significado fijado dentro de una institución didáctica dada para un cierto concepto. [...] un punto importante en la teorización usada es la consideración del significado de un concepto como una entidad compleja. Por ello conviene analizar separadamente sus componentes o “elementos de significado”, ya que los elementos específicos contemplados en los diversos textos podrían no ser los mismos o, incluso, alguno de ellos podría no aparecer en ninguno de los textos. Como resultado de lo anterior, se generaría un sesgo en el significado presentado a los alumnos. (Ortiz et al., 1996, p. 23)

En consecuencia, consideramos que el análisis de los textos es tópico fundamental que se debe tener en cuenta dentro de los aspectos del análisis de la enseñanza. Por ello, presentaremos en el siguiente capítulo, un análisis de las principales características de los textos. Dicho análisis se va centrar en aspectos relacionados con las unidades conceptuales que se presentaron en el tercer capítulo y que hacen parte del análisis de contenido de los tests de hipótesis.

SITUACIONES PROBLEMÁTICAS Y ENSEÑANZA

Butts (1980) clasifica los problemas que se proponen a los alumnos en cuatro tipos: ejercicios algorítmicos, problemas de aplicación, problemas de búsqueda y situaciones problemáticas. Mientras que los ejercicios que se realizan, en general, a mano o con calculadoras rudimentarias son de los tres primeros tipos, en clases donde se utilizan paquetes de análisis estadístico, hojas electrónicas o incluso calculadoras gráficas se pueden plantear situaciones problemáticas de análisis de datos.

Muchos textos de estadística adoptan una aproximación en cuanto a los problemas propuestos, para cuya solución basta identificar cual es la “receta” (léase algoritmo de solución) y el problema queda resuelto. Este tipo de problemas generan una visión equivocada de cómo deben aplicarse realmente las técnicas estadísticas. Es necesario proponer problemas que generen un ambiente didáctico más coherente con los principios de la enseñanza estadística y que contribuyan al enriquecimiento del conocimiento del estudiante. Se podría decir que en los textos tradicionales de estadística

se suelen presentar a los estudiantes situaciones problemáticas donde se refleja que:

- Se favorece el trabajo individual con respecto al trabajo en grupo.
- Cada problema se puede asociar claramente con una receta de solución.
- En general cada problema sólo tiene una respuesta o sólo se discute un método de solución.
- Se hace muy poco uso de recursos tecnológicos.
- La mayoría de los problemas se dejan para resolver individualmente en la casa y su solución se revisa en la clase.
- La comunicación de las ideas estadísticas se limita a lo relacionado con problemas del texto o a trabajos de investigación que no son desarrollados por los estudiantes.

De lo anterior se deriva que una de las maneras de aportar a la comunidad de educadores en estadística consiste en desarrollar procesos de innovación curricular en educación estadística que estén basados en las ideas constructivistas y que giren alrededor de la resolución de problemas. Desafortunadamente, en muchos de los materiales utilizados para el desarrollo curricular aún se percibe la ausencia de situaciones problemáticas que hagan énfasis en los siguientes aspectos:

- Que reflejen cómo “creemos que” se debe enseñar la estadística.
- Que muestren cuál es el papel que juega la resolución de problemas dentro de esa concepción.
- Que introduzcan un uso más generalizado de recursos tecnológicos tales como los computadores y las calculadoras gráficas.
- Que generen un mayor compromiso de los profesores para preparar las situaciones problemáticas y para ponerlas en práctica.
- Y como propósito último, que haga más eficiente el proceso de aprendizaje del estudiante, lo motive y mejore su comprensión.

Por otra parte, en la construcción del conocimiento estadístico también es importante considerar los análisis a nivel de la interacción social en la resolución de problemas. Cuando este tipo de análisis se realiza, los resultados deben llevar a la descripción del papel que juega, la interacción que tiene lugar entre los estudiantes con motivo de la resolución conjunta de una situación problemática. Cobb (1993) sugiere que en este tipo de análisis se debe tener en cuenta tanto la perspectiva psicológica —en la que se estudia el desarrollo matemático como un proceso individual de construcción concep-

tual —, como la perspectiva sociológica —en la que se enfatizan los aspectos sociales de la interacción de la actividad matemática—. Estos análisis deben aportar esquemas conceptuales y metodológicos que permitan observar, analizar y describir la manera como un grupo de estudiantes desarrollan individualmente y dan validez conjuntamente al conocimiento matemático relacionado con los conceptos objeto de estudio que construyen socialmente.

Uso de la tecnología

Las calculadoras y los computadores son un componente muy importante que se relaciona con el trabajo que se puede realizar en la enseñanza estadística. Proponer, por ejemplo, la realización de simulaciones y la exploración de comportamientos a través del computador y de las calculadoras es, además de un elemento motivante para el estudiante, un vehículo para analizar conceptos tales como la variabilidad o el efecto de diferentes tamaños de muestras sobre una estadística. Con la nueva tecnología tenemos acceso muy rápido a representaciones gráficas que hacen mucho más interesante y efectivo el análisis de datos para los estudiantes.

Con los computadores y también con las calculadoras gráficas tenemos la posibilidad de que el estudiante se concentre más en el significado de los resultados que en los procesos de cálculo. Sin embargo, la tecnología debe utilizarse con cuidado. La comprensión de muchos conceptos estadísticos debe estar bien estructurada antes de proceder a su utilización práctica en investigaciones. Buena parte del tiempo la deben invertir los estudiantes trabajando con los datos antes de lanzarse a interpretar los resultados que se pueden obtener cuando se usan computadores y paquetes estadísticos.

7. ANÁLISIS DE ALGUNOS ASPECTOS DE LA ENSEÑANZA EN LOS TESTS DE HIPÓTESIS

INTRODUCCIÓN

En este capítulo se presentan dos secciones. En la primera se mencionan algunas de las características de los textos que se han utilizado durante los últimos semestres en los cursos de Estadística 1 y Estadística 2. En particular, se comenta acerca de algunos tópicos de estadística que están relacionados con las contrastes de hipótesis. En la segunda sección, se hace alusión a algunas características que consideramos que son relevantes en relación con los problemas de estadística que se presentan en la enseñanza. Además, en esta sección también se incluyen algunos comentarios acerca del papel mismo del profesor en la enseñanza.

ANÁLISIS DE LOS TEXTOS

Textos utilizados en Estadística 1 y Estadística 2

Para el primer semestre de 1997 el texto principal del curso Estadística 2 fue *Elementary Statistics*, de R. Johnson. Sin embargo, durante los últimos tres semestres también se utilizaron otros textos tales como *Estadística para las ciencias sociales* de R. Runyon y A. Haber y *Estadística y Sociedad. Conceptos y herramientas básicas del método estadístico* de Fernández et al. Durante el segundo semestre de 1996 también se utilizó en el curso de Estadística 1 el primer texto mencionado en dos de las seis secciones y en las otra cuatro secciones se utilizó el texto *Matemáticas, Azar, Sociedad. Introducción a los conceptos básicos de estadística* de Perry et al. En los que sigue vamos a hacer referencia a los textos anteriores. Para ello los hemos numerado así:

- [1]: Johnson, R. (1996) *Elementary Statistics*. Washington: Duxbury Press. (séptima edición, aún no ha sido traducida al español).
- [2]: Runyon, R & Haber, A. (1987). *Estadística para las ciencias sociales*. México: Addison-Wesley Iberoamericana (traducción al español de la cuarta edición publicada en inglés).
- [3]: Fernández, F. (1996). *Estadística y Sociedad. Conceptos y herramientas básicas del método estadístico*. Bogotá: “una empresa docente” (primera edición).

- [4]: Perry, P. et al. (1996). *Matemáticas, Azar, Sociedad. Introducción a los conceptos básicos de estadística*. Bogotá: “una empresa docente” (segunda edición).

Características del contenido

El análisis de estos cuatro textos da una primera visión del enfoque pedagógico seguido en los cursos de Estadística 1 y Estadística 2. Los textos [3] y [4] están ligados por una misma propuesta de diseño curricular (ver por ejemplo, Perry et al. 1995), en este sentido el texto [4] fue diseñado para ser visto en el curso Estadística 1 y el texto [3] para ser visto en el curso Estadística 2. Por otra parte, los textos [1] y [2] están diseñados para ser vistos a lo largo de dos semestres. En [1] se presenta, en su séptima edición, una intención explícita de desarrollar los conceptos haciendo más énfasis en la construcción crítica de habilidades de pensamiento que en lo computacional. En particular, en el capítulo 8 “Introducción a la inferencia estadística” se presenta un flujo muy natural de ideas que van desde las distribuciones de muestreo a la estimación y luego a los tests de hipótesis. El texto [2] es un libro más antiguo y contrasta con [1] por presentar un enfoque “más tradicional” de la enseñanza de la estadística y un mayor énfasis en el manejo de fórmulas y en la computación de resultados que en su interpretación.

En esta sección queremos hacer referencia a la forma de presentación de algunos de los temas a los que nos hemos referido en el análisis de contenido. La intención es caracterizar la manera como estos temas se presentan y cómo pensamos que dicha presentación influye en la enseñanza.

Población, muestra y parámetro

Estos conceptos, por estar presentes de manera explícita o implícita en todo lo que se hace en estadística, deberían recibir una gran atención en la mayoría de los textos. Sin embargo, nuestra opinión es que estos términos no son definidos en los textos de una manera suficientemente precisa y coherente. Por ejemplo, en [1] y [2] en la definición del término población no se hace explícita la diferenciación entre población de estudio y población de datos.

En el texto [2], se define población como el conjunto completo de individuos, objetos o mediciones que tienen alguna característica común observable, o como el conjunto teórico de observaciones potenciales. Mientras que en el texto [4], se define población de investigación, o simplemente la población de estudio, como el conjunto de todos los entes a los cuales se pueden aplicar las conclusiones obtenidas a través de la predicción, la estimación o la verificación de una hipótesis. Además en [4], luego de definir población de investigación, se define la población de datos o simplemente la población como el conjunto de todas las mediciones que es posible obte-

ner a partir de observar una cierta característica en cada uno de los elementos de la población de estudio.

Por supuesto, en todos los casos referidos anteriormente los comentarios que se acaban de mencionar acerca de la definición de la población, también se aplican a la definición de muestra de estudio y a la de muestra de datos.

En cuanto a la definición de parámetro, en todos los textos analizados dicha definición se da en términos lingüísticos muy similares. En todos ellos, se hace alusión a la convención de utilizar letras griegas para representar los parámetros, sin embargo, no se hace caso a esta convención cuando se trata de la proporción, aquí la diferencia se establece entre una “p” minúscula (texto [2]) o una “ \hat{p} ” (con el gorro) para denotar el estimador (textos [1] y [3]) y una “P” mayúscula para denotar el parámetro.

Datos, variables y escalas

Los tipos de datos o las variables que se manejan en estadística suelen clasificarse de acuerdo con diferentes criterios. Uno de estos criterios se da en términos de las escalas de medición con base en la cual se expresan las mediciones. Al respecto, hay un tipo de clasificación que se presenta con frecuencia pues parece adaptarse más apropiadamente a los tipos de datos que se obtienen en las ciencias sociales. Se puede expresar de manera breve en términos de la secuencia: nominal, ordinal, discreto y continuo. Por ejemplo, en el texto [2] (p. 6 y pp. 20-29) se hace precisamente la presentación de estas nociones, siguiendo esta línea de argumentación. Por ejemplo se dice:

De hecho, hay tres formas fundamentales distintas de utilizar los números: 1. Para designar (números nominales). 2. Para representar la posición en una serie (números ordinales). 3. Para representar una cantidad (números cardinales)”. (p. 21)

Otro de los criterios tiene que ver con características de diseño en el sentido experimental, se habla entonces de variables dependientes o independientes o de variables de control y experimentales.

Por otra parte, en el texto [1] (pp. 14 -18) cuando se consideran las nociones de datos o de variables, no se habla de estos criterios de clasificación en un sólo capítulo del libro, sino que se hace en forma muy dispersa a lo largo del libro. Aunque en el texto [3] tampoco se hace alusión a estos criterios de clasificación, en el texto [4], si se dedica todo un capítulo a hablar del tema de variables (pp. 41-51).

Vemos, con base en los comentarios anteriores, las diferencias en el énfasis que se da, en cada texto, a estos conceptos. Esto lógicamente trasciende en la forma como el profesor da explicaciones sobre estas nociones, y en la forma como el estudiante asimila dichos conceptos.

Medidas de tendencia central, variabilidad y correlación

Estas nociones, están presentes de manera explícita, en la mayoría de los textos de estadística. En [1] y [2] se presentan uno tras otro cada uno de estos temas, sin embargo, en [4], no se considera el tema de correlación cuantitativa. Por otra parte, en el texto [3] se considera de manera consecutiva la parte descriptiva y luego la parte inferencial con un tratamiento en el que se presenta el coeficiente de correlación a través de un proceso de construcción muy didáctico (Fernández y Monroy, 1995). Esta construcción utiliza los conceptos de localización, de escala y de estandarización para llegar a la expresión tradicional de la correlación de Pearson.

En el texto [4] se introduce el estudio de estos temas con una aproximación intuitiva y empírica. Por ejemplo, en el capítulo que trata sobre medidas de tendencia central (pp. 92-103) se introduce, a través de un diálogo, el concepto de media aritmética; la aproximación presentada enfatiza en la construcción conceptual del promedio y refleja claramente una posición constructivista ante la enseñanza de la estadística. Esta posición también se puede percibir claramente en la forma como se abordan la definición de algunas medidas de variabilidad tales como el rango y la desviación estándar (pp. 110- 124).

Nivel de significación y p-valor

En el análisis de contenido y de aprendizaje fue considerado la comprensión de estos dos conceptos como muy relevante para el proceso de tests de hipótesis. Ahora vamos a mencionar la forma como aparecen y como se utilizan en los textos que hemos considerado.

Los conceptos de nivel de significación y p-valor se constituyen en uno de los focos principales de investigación en este proyecto. En general, por lo que se puede ver en el texto [1], parece que hay una tendencia en los nuevos textos a hacer un uso más extensivo del concepto de p-valor. También parece, por lo que se puede ver en los demás textos ([2], [3] y [4]), que era menor la importancia que se le daba al uso de este concepto en años anteriores. Aunque la veracidad de lo dicho anteriormente puede ser cuestionada si se consideran otros contextos de enseñanza, o con base en referencias de otros textos de enseñanza, al menos si es cierto en el contexto educativo donde realizamos esta investigación.

Para empezar, en el texto [2] (p. 220) se hace una descripción en seis pasos del procedimiento para realizar un test de hipótesis. Dicha descripción corresponde con lo que en el texto [1] se llama aproximación clásica (p. 403). En la ‘aproximación clásica’ es muy claro que no se utiliza el concepto de p-valor. En realidad esta es la aproximación que se presenta en todos los textos que hemos considerado, excepto en el texto [1]. Aunque en el texto [2] no se define formalmente el concepto de p-valor, si se utiliza de manera implícita en algunos pocos ejercicios (por ejemplo en el ejercicio nueve, p. 212). En general, en los textos [2], [3] y [4] los ejercicios y pro-

blemas de tests de hipótesis que se plantean al estudiante son para resolverlos siguiendo la aproximación clásica. Igualmente, en [2], [3] y [4] se ilustra por medio de ejemplos resueltos la aplicación de la aproximación clásica.

Sin embargo, sí se perciben algunas diferencias con respecto a los problemas que se presentan en [3] y [4] comparados con los que se presentan en [2]. En los dos primeros se plantea, con mucha más frecuencia, problemas donde no sólo se debe aplicar el algoritmo de la aproximación clásica, sino donde también se debe razonar sobre los resultados mismos que dan las estadísticas de prueba o los estimadores en términos de los niveles de significación (ver por ejemplo, en el texto [4], los problemas 9 y 11 de la p. 282 o en el texto [3] los problemas 11 y 12 de las pp. 99-107).

Con respecto al uso del nivel de significación, una de las cuestiones que ya hemos mencionado que puede tener relevancia en el diseño de situaciones problemáticas es la fijación del mismo. En la actualidad este punto es objeto de muchas discusiones. Por ejemplo, se puede dar cuenta de discusiones al respecto que se han dado en la lista edstat-l@jse.stat.ncsu.edu¹ sobre lo que llamaron “la mitología del valor $\alpha = 5\%$ ” (ver por ejemplo, los mensajes en febrero del 97 de Burrill y Rubin entre otros). Con respecto a este asunto, al hacer un pequeño muestreo, vimos que mientras en el texto [2] se puede decir que en alrededor del 60% de los ejercicios se fija de antemano el nivel de significación (usualmente se fijan 5% o 1%), en los textos [3] y [4], se tiende a no fijarlo de antemano. Por otra parte, en el texto [1], a pesar de utilizar con mucho énfasis la noción de p-valor también se suele establecer de antemano un nivel de significación. Esto es explicable en parte, por la necesidad que se tiene de hacer que el problema tenga una respuesta precisa (se rechaza o no se rechaza).

Finalmente y volviendo al texto [1], en este se hace una presentación formal y detallada de dos algoritmos que pueden ser utilizados para solucionar los problemas de tests de hipótesis. El autor los llama ‘aproximación clásica’ y ‘aproximación al p-valor’. Esencialmente la diferencia entre ambas aproximaciones es el tipo de comparación que se establece para tomar la decisión de rechazo o no de la hipótesis nula: mientras que en la primera se comparan valores del dominio de un modelo de distribución, en la segunda se comparan probabilidades de la distribución.

Formulación de hipótesis

Este es uno de los aspectos que se debería manejar con más cuidado en los textos, pues detrás del tratamiento que se le da a la formulación de las hipótesis está presente la lógica del contraste.

A la hipótesis nula usualmente se le atribuye el sentido de ser la hipótesis donde se expresa la nulidad o algo así como “no han cambiado las

1. Lista de discusión en Internet sobre temas de enseñanza de la estadística.

cosas”. Por ejemplo, en el texto [3] (p. 242 a 254) se compara la lógica de la formulación de hipótesis con el procedimiento que se sigue en un juicio para juzgar a un acusado. Se utiliza en un sentido metafórico, la comparación entre la hipótesis nula y la hipótesis de inocencia o “el principio de inocencia” de un acusado.

En el texto [2] (pp. 206-207) se utiliza un enfoque en el que se menciona la noción de prueba indirecta. Se refieren a la prueba que en la lógica matemática corresponde a la prueba por contradicción o reducción al absurdo. Sin embargo, en este caso es una prueba indirecta en un sentido inferencial. Por ejemplo, en dicho texto dice:

[...] por el contrario, puesto que la hipótesis alternativa no puede ser probada ni rechazada directamente, no podremos probar nunca la hipótesis nula rechazando la hipótesis alternativa. [...] en la práctica, muchos investigadores usan indistintamente las expresiones «fracasar en rechazar la hipótesis nula» y «aceptar la hipótesis nula». (p. 206)

Desde nuestro punto de vista, pensamos que no es fácil para los estudiantes comprender en toda su complejidad la lógica que hay detrás de la formulación de las hipótesis. Por ello se cae, casi que irremediamente, en aproximaciones como las que se presentan en el texto [1] que simplifican y mecanizan y el razonamiento que debería estar presente detrás de la formulación. Esto se puede ilustrar en el texto [1] (pp. 390 y 407) en un cuadro donde se resumen las frases más comunes que aparecen en las representaciones verbales de una situación de test de hipótesis y la forma como se pueden traducir a la formulación del test de hipótesis en términos simbólicos (ver Tabla N° 6).

Tabla N° 6. Frases más comunes y sus negaciones en la formulación de hipótesis.

Ho: (\geq)	Ha: ($<$)	Ho: (\leq)	Ha: ($>$)	Ho: ($=$)	Ha: (\neq)
al menos	menor que	a lo más	más que	es	no es
no menos que	menor que	no más que	más que	no diferente de	diferente de
no menor que	menor que	no tan grande que	más grande que	igual a	no igual a

A nuestro modo de ver, de aquí se deriva indirectamente una regla de acción que en la práctica de la formulación de hipótesis se traduce en la frase: “siempre debe haber un igual en la hipótesis nula”. Esta es una regla de ac-

ción que posibilita el que los estudiantes resuelvan los problemas, pero no civa en términos de la comprensión de la lógica que hay detrás del contraste.

Otro punto sobre el cual encontramos necesario hacer algunos comentarios, es la poca presencia en prácticamente todos los textos, de la formulación de hipótesis alternativas simples. La Tabla N° 1, refleja implícitamente este hecho. Las consecuencias de esto son una adherencia mucho más marcada por una posición “fisheriana” hacia el test de hipótesis que hacia la posición de Neyman-Pearson de ver un test de hipótesis no como poniendo a prueba una de las hipótesis (la nula), sino como un contraste de dos hipótesis.

Distribuciones

En esta sección nos vamos a referir a algunas de las características de los textos en cuanto a la presentación de los conceptos de distribución muestral y modelos de distribución. Con respecto a los modelos de distribución miramos en particular la forma como se presentan las ideas de valores críticos y regiones de rechazo.

Para empezar, los textos [1] y [3] se caracterizan por introducir el tema de distribución muestral (pp. 328 y 206 respectivamente) por medio de preguntas. Por ejemplo, en [1] luego de decir que se ha extraído una muestra de la población, se pregunta: ¿usted espera que el valor de la media muestral sea exactamente igual al valor de la media de la población? y en [3], se pregunta: ¿qué tan segura es la inferencia que se puede hacer con base en la información que arroja una sola muestra? Aunque la pregunta planteada en [3] es más general, en ambos textos ([1] y [3]) la argumentación introductoria desemboca en la presentación de ejemplos hipotéticos de poblaciones pequeñas en las que se simula el proceso de extracción de muestras. En este sentido, el texto [2] presenta algunas diferencias con respecto a [1] y [3]. En primer lugar, se presenta la definición formal al principio de la argumentación introductoria y no al final, en segundo lugar, prácticamente no se utilizan representaciones gráficas y tabulares del proceso de construcción de la distribución muestral y en tercer lugar, no se establece una conexión inmediata con el teorema de límite central para la media muestral.

En cuanto al texto [4], por ser esta continuación del texto [3] en la propuesta de diseño curricular que se había planteado inicialmente, se consideran muy pocas discusiones alrededor del concepto de distribución muestral. En todo caso, para mostrar algunas de las diferencias entre las distribuciones t-student y normal, se ilustra, a través de un ejemplo de construcción de una distribución muestral, la diferencia en la variabilidad (pp. 29-32).

Con respecto a las formas de representación gráficas y tabulares de las distribuciones, hay que señalar que la introducción del uso del p-valor, motiva una diferencia en la presentación de algunas cuestiones en los textos. Por ejemplo, las tablas que se presentan en el texto [1] presentan aspectos que no se muestran en los otros textos. Destacamos los siguientes:

- Se presentan de manera formal tablas de p-valores para las distribuciones normal y t-student.
- Se utilizan de manera explícita los símbolos α o p-valor en las representaciones gráficas de las tablas de distribución.
- Presenta una tabla donde se puede mirar directamente la significación del resultado de una correlación muestral.
- Presenta una tabla que permite estimar de manera directa intervalos de confianza del 95% para el coeficiente de correlación de una población.
- Presenta una pequeña extracción con los valores más utilizados para situaciones de tests unilaterales y bilaterales con sus respectivas representaciones gráficas (p. 761).

PROBLEMAS DE ESTADÍSTICA Y SU ENSEÑANZA

Con respecto al papel de los profesores en la enseñanza de la estadística vamos a hacer algunos comentarios que se basan en las experiencias que se han compartido con los diferentes profesores que han dictado los cursos de Estadística 1 y Estadística 2, desde 1989 a 1996. La interacción con estos profesores se generó principalmente en las reuniones de coordinación que se desarrollaron a lo largo de este periodo y en algunas charlas informales.

Hemos organizado los comentarios tratando de responder a varias de las preguntas que se relacionan con la forma como se han manejado los problemas de estadística que se les proponen a los estudiantes y a la forma como se manejan las clases de estadística. Queremos de esta manera reflejar como es la situación con respecto a cuestiones que pensamos favorecen o desfavorecen el cómo creemos que se debe enseñar estadística.

¿Se favorece el trabajo individual con respecto al trabajo en grupo?

Al respecto pensamos que hay que ganar experiencia para poder dirigir discusiones cuando se trabaja con los estudiantes dentro de una atmósfera abierta a la experimentación. Predominan los contextos que favorecen el individualismo, pero pensamos que se pueden cambiar por situaciones donde los estudiantes, formados en pequeños grupos, intercambien sus ideas.

¿La mayoría de los problemas se dejan para resolver individualmente en la casa y su solución se revisa en la clase?

Los profesores en general están sujetos a cumplir el programa cueste lo que cueste. Usualmente, se limitan a corregir ellos mismos las tareas asignadas en el programa para la casa o a pasar al tablero a un estudiante para revisarlas. Luego, pasan rápidamente al próximo tópico y se repite nuevamente

esta rutina. Sin embargo, cuando se piensa que la estructura de una clase de estadística no se debe ceñir de manera estricta a una programación preestablecida, se deja pie para una mayor libertad de acción. Por ejemplo, una tarea que inicialmente se había previsto para una clase podría tomar tres días para que los estudiantes realmente tengan tiempo de experimentar y desarrollar “correctamente” un problema. Muchas veces se presentan ideas que son sugeridas por los estudiantes, que vale la pena que sean consideradas en la mitad de una clase, a pesar de que el tiempo se pueda agotar y no se alcancen a asignar las tareas del día siguiente. Es difícil que un profesor este dispuesto a aceptar que esto suceda.

¿En general cada problema sólo tiene una respuesta o sólo se discute un método de solución?

Parece que es problema de todos los profesores, el que tengan dificultades para transmitir, manejar o generar situaciones didácticas donde “las matemáticas” que se trabajen no necesariamente tengan una única respuesta. Se reconoce la dificultad para construir situaciones donde se pueda hacer explícito el que haya varias respuestas dependiendo de cómo se interprete el problema. Y aunque la mayoría de profesores parece que comprenden los procesos de enseñanza que se deben llevar a cabo, no se involucran realmente con el espíritu de los materiales y de las técnicas del desarrollo curricular. Con mucha frecuencia se omiten algunos puntos de análisis y se enfoca el desarrollo del curso mucho más en la maestría del manejo de herramientas. Por ejemplo, en el diseño de proyectos de investigación que se proponen los estudiantes, usualmente no se reconocía que un mismo conjunto de herramientas básicas puede ser aplicado con diferentes niveles de sofisticación y a diferentes conjuntos de datos, o que diferentes maneras de analizar un mismo conjunto de datos, pueden llevar a revelar diferentes conceptos.

¿Se hace muy poco uso de recursos tecnológicos?

En los cursos Estadística 1 y Estadística 2, los profesores y los estudiantes han utilizado en semestres anteriores recursos tecnológicos tales como computadores y más recientemente calculadoras gráficas.

En lo que se refiere a computadores se ha utilizado el programa de análisis estadístico Stat View (versión 4) y la hoja de cálculo Excel (versión 4 para Macintosh). Y en cuanto a la calculadora gráfica se utilizó la TI-85 de Texas Instruments. A continuación comentamos brevemente algunas de las principales características de tipo didáctico relacionadas con cada uno de estos.

Stat View. Fue utilizado en los cursos de Estadística 1 y Estadística 2 a finales de los años 80 y comienzos de los 90. De hecho en la segunda edición de Perry et al. (1989, pp. 359-379), se incluye un capítulo que ilustra ele-

mentos básicos de su manejo. Su utilización se dio principalmente en el contexto de proyectos de investigación que fueron realizados por estudiantes. Un ejemplo con las características y el contenido de este tipo de proyectos se encuentra en Jaramillo (1989).

Excel. Ha sido principalmente utilizado en el curso de Estadística 2, como apoyo para la realización de proyectos similares a los mencionados en el párrafo anterior. Aunque contiene un módulo especial para análisis estadístico, su utilización también se puede enfocar para hacer énfasis en la definición de procedimientos. Por ejemplo, se pueden proponer tareas en donde los estudiantes deban definir concretamente el proceso de un test de hipótesis para un conjunto de datos y hacer que ellos mismos se ideen la forma en que deben resumir y presentar los resultados.

Calculadora gráfica TI-85. Su utilización se introdujo en el segundo semestre de 1994 en el curso de Estadística 2, para introducir el tema de correlación (Fernández y Monroy, 1995). Luego se extendió su utilización para el manejo de otros temas. En particular, su utilización se tuvo en cuenta en el diseño, desarrollo y evaluación de las tres situaciones problemáticas de tests de hipótesis que se analizan en esta investigación.

¿Cómo son las actitudes y la motivación de los estudiantes hacia los problemas que se les presentan?

La experiencia nos dice que es difícil lograr que ellos “se apropien” verdaderamente de un problema estadístico. Aunque creemos que el gusto y las actitudes por la estadística pueden cambiar en la medida en que las situaciones que se presentan al estudiante sean de su interés, pensamos por una parte, que si las situaciones que se trabajan están enfocadas más en el interés de ellos, los estudiantes pueden tener la creencia de que no son “verdaderas matemáticas” creencia que en algunos casos puede constituirse en el mayor obstáculo. Y por otra parte, vemos que los estudiantes encuentran frecuentemente dificultades para comunicar sus ideas y para la realización de análisis. Por ejemplo, dado un problema, ellos pueden hacer un test de hipótesis y llegar a una respuesta, pero posiblemente no comprenden, en términos del contexto del problema, lo que se les está preguntando.

Por otra parte, el progreso de los estudiantes muchas veces se gradúa con base en las respuestas correctas que ellos dan a pruebas de evaluación que no son coherentes con el aprendizaje que realmente deseamos que ocurra en ellos. Es necesario proponer nuevas formas de evaluación consistentes con nuestras metas de aprendizaje y de desarrollo de actitudes. Así pues, esta fue una de las metas que nos propusimos al hacer el diseño de las tres situaciones problemáticas de estadística. En los capítulos que siguen se presentan los detalles del diseño, de la aplicación y de la evaluación de dichas situaciones problemáticas.

8. SITUACIÓN PROBLEMÁTICA N° 1: INFERENCIA ACERCA DE LA PROPORCIÓN DE UNA POBLACIÓN

INTRODUCCIÓN

En esta situación problemática se trabajó el tema de tests de hipótesis para inferir acerca de la proporción en una población. El diseño contempló la elaboración de dos problemas: uno que se pudiera resolver con base en la aplicación de la distribución binomial o con la aproximación normal de dicha distribución, y el segundo, que se pudiera resolver, sólo con la aplicación de la distribución binomial.

Para llegar a la concreción del diseño se hizo un análisis didáctico que contempló tres dimensiones: por un lado, se realizó un análisis del contenido estadístico del tema que siguió el esquema planteado en los capítulos 2 y 3. Luego, se hizo el análisis de algunos de los aspectos del aprendizaje. En particular, se eligió concretamente un tipo de situación problemática y se miró en ésta, los conceptos y procedimientos que se ponían en juego, se hicieron algunas conjeturas acerca de los posibles errores que podrían cometer los estudiantes al resolver el problema y también se miraron algunos aspectos relacionados con el lenguaje estadístico utilizado. Este último, se hizo con base en el análisis del uso de representaciones externas y sus formas de traducción. Por último y con base principalmente en el análisis del texto que debían leer los estudiantes antes de enfrentarse al problema propuesto, se miró la dimensión de la enseñanza.

La forma como se aplicó la situación problemática se puede resumir en tres pasos. Primero, los estudiantes resolvieron la situación de manera individual y la entregaron por escrito. Segundo, los profesores la revisaron, identificaron y luego señalaron los errores de los estudiantes para planear la sesión de trabajo en grupo. Tercero, los estudiantes se reunieron de nuevo en la clase, trabajaron en grupo y su tarea consistió en la explicación y en la corrección de los errores que se les habían señalado. Además, se miró con más detalle la forma como uno de los grupos de tres estudiantes explicó y corrigió los errores.

ANÁLISIS DEL CONTENIDO ESTADÍSTICO

En esta sección solamente se hace alusión a los conceptos que están referidos en cada una de las unidades de análisis conceptuales propuestas en los capítulos 2 y 3. Dado que los posibles conceptos y procedimientos que realmente se puedan utilizar en la acción dependen de la situación y de la

formulación del problema que se proponga finalmente a los estudiantes, esta especificación se hará en la sección que presenta los aspectos del análisis del aprendizaje.

A. Datos

Variables

Los problemas de estadística que se relacionan con la inferencia acerca de una proporción consideran de manera implícita o explícita lo que en la literatura se denomina como experimentos aleatorios de Bernoulli. Un experimento de Bernoulli sólo tiene dos resultados posibles que suelen denominarse con el nombre de éxito o fracaso. Usualmente se asocia el valor 1 al éxito y el valor 0 al fracaso. Entonces, el valor de una variable binomial se obtiene cuando se hace el conteo del número de éxitos que se obtienen al realizar n veces un experimento de Bernoulli.

Población

El tamaño de la población que normativamente se debe asociar a una experiencia de tipo binomial es infinito. Esto se debe a los supuestos de independencia y de permanencia constante de la probabilidad de éxito que caracterizan un experimento binomial. Sin embargo, aunque en la práctica la mayoría de poblaciones consideradas son finitas, el tamaño de la muestra extraída de la población usualmente será tan pequeño comparado con el de la población, que para propósitos prácticos podrá considerarse que la población es infinita.

Muestra

Dado que en la práctica la mayoría de poblaciones son finitas, para obtener una muestra que satisfaga el supuesto de la permanencia constante del valor de la probabilidad de éxito, se debe hacer un muestreo con sustitución. En todo caso, cuando suceda que el tamaño de la muestra es relativamente grande comparado con el de la población se debe considerar según la teoría de inferencia estadística, un factor de corrección para poblaciones finitas.

Nivel de significación

Cuando las muestras son muy pequeñas es posible que no tenga sentido establecer a priori niveles de significación tradicionales tales como 5% o 1%. Esto es debido a que la distribución binomial no es continua y por lo tanto no se pueden establecer niveles de significación sobre un espectro de valores que sea continuo.

Medidas de resumen estadístico

Hay tres formas como se pueden establecer medidas de resumen asociadas con un experiencia de tipo binomial para reportar el número de éxitos. Una,

en términos de porcentajes; otra, en términos de proporciones y la otra en términos de frecuencias absolutas. Formalmente, la proporción se puede definir como una media aritmética (suma de los unos y ceros de los éxitos y fracasos de los experimentos de Bernoulli dividido por el tamaño de la muestra) y por tanto se podría considerar como una medida de tendencia central.

B. Formulación

Parámetros

Aunque como se mencionó anteriormente, hay tres maneras de resumir estadísticamente los resultados asociados a una experiencia binomial, la proporción del número de éxitos en una población es la forma usual como se establece el parámetro de estudio. De otro lado, también la proporción poblacional se identifica con la probabilidad de éxito de la variable dicotoma considerada en un experimento de Bernoulli.

Hipótesis de investigación

Con relación a la forma en que se hace referencia a la hipótesis de investigación, en general, se tendrá una sospecha concreta con respecto al valor de la proporción en la población. Sin embargo, también se pueden presentar casos en donde no se menciona de manera explícita el valor del parámetro. Por ejemplo, uno de los casos más conocidos es el de la moneda que se sospecha que está cargada. En este caso, como se explica en la próxima sección, se recurre al “Principio de indiferencia” para el planteamiento de las hipótesis estadísticas.

Hipótesis estadísticas

El “Principio de indiferencia” predica atribuir las misma probabilidad (0.5) de ocurrencia a los dos posibles resultados de la variable de Bernoulli. Por esto, en el ejemplo donde se considera una secuencia de lanzamientos de una moneda, el planteamiento de las hipótesis estadísticas con las cuales se quiere determinar si la moneda está o no cargada, debe asumir, bajo la hipótesis nula, el “Principio de indiferencia”, es decir, suponer que la moneda no está cargada. Por ejemplo, si no se tiene sospechas de hacia que lado está cargada la moneda, el planteamiento debería ser de la forma $H_0: p = 0.5$ contra la alternativa $H_a: p \neq 0.5$.

Lógica del contraste

Además de la lógica general que hay detrás de los contrastes de hipótesis, sólo se presentan, como casos particulares, los problemas donde es necesario recurrir al “Principio de indiferencia”. Con la diferencia de que en estos casos “la nulidad” no se debe interpretar como un cero sino como un valor de 0.5.

C. Herramientas

Estimadores

En la aproximación a la distribución normal se pueden considerar como estimadores tanto el total de éxitos (representado simbólicamente como “x”), como la proporción muestral (es decir, “x/n”, representado simbólicamente como “ \hat{p} ” y donde n representa el tamaño de la muestra). Para el caso en el que no se quiera o no sea apropiado realizar la aproximación normal, no es necesario calcular un estadístico de prueba. En este caso estaremos trabajando directamente con la distribución binomial.

Varianza del estimador

Al utilizar la aproximación normal de la distribución binomial, se pueden presentar dos casos. El primero, ocurre cuando se considera directamente el número de éxitos, caso en el cual el error estándar asociado es \sqrt{npq} , donde “p” representa el valor de la proporción en la población, “q” es “1 - p” y “n” es el tamaño de la muestra). El segundo caso ocurre cuando se considera la proporción muestral como estimador, en este caso el error

estándar de la proporción se estima con $\sqrt{\frac{pq}{n}}$.

Estadísticos para tests de hipótesis

Si se trabaja con la distribución binomial no es necesario calcular un error estándar, pues la significación del resultado muestral se establece directamente con base en la distribución que se debe elegir según el supuesto que se tenga en la hipótesis nula. Cuando no se utiliza la distribución binomial, el estadístico de prueba está dado por el cociente entre la diferencia del total de éxitos obtenido y valor el esperado bajo la hipótesis nula o por la diferencia entre la proporción muestral y el valor de la proporción bajo la hipótesis nula, dividido por el respectivo error estándar, según el caso considerado. En términos simbólicos estas expresiones son respectivamente:

$$\frac{x - np}{\sqrt{npq}} \quad \text{y} \quad \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

D. Distribuciones

El tests de hipótesis para una proporción está basada en una distribución binomial y por lo tanto cuando esta se aplica se estará ante un test de tipo no paramétrico. Esto quiere decir que se estará ante un test donde no es necesario estimar otros parámetros diferentes al de la proporción misma (como por ejemplo la desviación estándar) para determinar completamente

la distribución. Tampoco es necesario suponer que la población de la cual se extrae la muestra deba o no seguir una distribución específica.

Distribución muestral de proporciones

La distribución muestral de proporciones queda completamente determinada con la distribución binomial. Esta distribución, al igual que en los casos de las otras dos situaciones que fueron trabajadas en esta investigación, se puede aproximar por medio de la distribución normal si se satisfacen condiciones relacionadas con el tamaño de la muestra y el valor que tenga la probabilidad de éxito.

Distribución del estimador

Los modelos de distribución de probabilidad asociados con los estimadores de proporción o del total de éxitos son el binomial y el normal. En teoría siempre se debería utilizar el modelo binomial. Sin embargo, es usual que por dificultades técnicas no se pueda utilizar la distribución binomial, lo cual ocurre normalmente cuando el tamaño de la muestra es muy grande.

Distribución del estadístico de prueba

Realmente sólo se considera la utilización de un estadístico de prueba cuando no se puede utilizar la distribución binomial. Si este es el caso, pueden utilizarse los estadísticos de prueba mencionados en el tercer apartado de la sección de herramientas. También se puede entonces, utilizar la distribución normal si se satisfacen criterios de aproximación asintótica.

E. Análisis e interpretación de resultados

Los conceptos que se manejan a nivel de esta unidad de análisis son generales y aplicables a las otras situaciones problemáticas que se van a considerar. De todas formas haremos una breve referencia a los conceptos que se ponen en juego en esta unidad.

Nivel de significación estadística

Se define con base en la distribución teórica que se utilice y teniendo en cuenta el establecimiento de la hipótesis nula. Formalmente se puede expresar como $\alpha = P(\text{Rechazar } H_0 / H_0 \text{ es cierta})$. En este caso $P(I)$ sería la función de densidad de probabilidad condicionada por la hipótesis nula. Se debe entonces calcular con base en una distribución binomial o una normal.

Tipos de error

Hay dos tipos de error, el de tipo I o α que consiste en rechazar la hipótesis nula cuando en realidad es verdadera, y el de tipo II o β que consiste en no rechazar la H_0 cuando en realidad es falsa. Entre menor sea el nivel de

rechazo, menor será la probabilidad de cometer un error de tipo I y mayor será la probabilidad de cometer un error de tipo II y viceversa.

P-valor

Dado el resultado muestral, el p-valor es el mínimo nivel de significación que es necesario establecer para empezar a rechazar la hipótesis nula.

Dependiendo del planteamiento de las hipótesis tiene una definición que varía según que el test sea de dos colas, o sea unilateral a la izquierda o a la derecha.

ANÁLISIS DE APRENDIZAJE

Esta sección se divide en tres subsecciones. En la primera se describen algunas de las características de los tipos de situaciones problemáticas que consideran inferencias acerca de la proporción de una población. En la segunda se presentan las dos situaciones problemáticas que se propusieron a manera de ensayo piloto, la manera como se aplicó y algunos comentarios relacionados con la evaluación de las mismas. Finalmente, en la tercera subsección, se presenta la propuesta definitiva, los conceptos y procedimientos que se ponen en juego en dicha propuesta y los errores que pensamos —a priori— que se iban a identificar en los estudiantes y algunos comentarios acerca del uso de representaciones externas relacionadas con el concepto de nivel de significación.

Situaciones problemáticas: caracterización del tipo de problemas

Como una primera aproximación para analizar los fenómenos que usualmente ponen en juego el tema que nos ocupa, se hizo la recopilación de una serie de problemas de inferencia acerca de la proporción de una población que aparecen en la práctica. Con base en dicha recopilación se miraron algunos los constructos hipotéticos utilizados en cada problema, es decir, la idea general del fenómeno a partir del cual se planteaba el problema.

Como resultado de la revisión hecha, y sin tener en cuenta la frecuencia con la que se presentó cada constructo, se concluyó que se utilizan estimaciones de proporciones en campos tales como: mercadeo, juegos de apuesta, demografía, psicología, industria, ciencia política y salud, entre otros. Algunos ejemplos particulares relacionados con cada uno de los campos mencionados son:

- Proporción de usuarios de un cierto artículo (mercadeo).
- Probabilidad de ganancia en una ruleta (juegos de apuestas).
- Proporción de la presencia de una cierta característica demográfica (demografía).

- Proporción de personas que reaccionan favorablemente ante cierto estímulo (psicología).
- Proporción de artículos defectuosos en un lote de producción (industria).
- Proporción de votantes por un candidato político (ciencia política).
- Proporción de personas contagiadas por el SIDA (salud).

Por otra parte, la revisión de los tipos de problemas que usualmente se presentaban se clasificó en dos grandes tipos: problemas de estimación de una proporción (sin tener conocimientos a priori de su posible valor) y de la revisión de la ocurrencia de una cierta proporción.

Problemas de estimación de una proporción. Este tipo de problema se le plantea con mucha frecuencia a compañías de consultoría. Por ejemplo, los muestreos de opinión que se publican en la prensa que reflejan las posibilidades de que un candidato dado gane las elecciones, son problemas donde se estima la proporción de personas que estarían dispuestas a votar por el candidato. En general este tipo de estimaciones suele utilizar dos conceptos estadísticos: margen de error y nivel de confianza. En este tipo de casos indirectamente se puede generar un problema de decisión para el candidato: seguir o no en la candidatura.

Problemas de revisión de la ocurrencia de una cierta proporción. En esta categoría se consideran los problemas donde se plantea la necesidad de controlar con determinada frecuencia si un proceso, en el que se sabe que usualmente ocurre un fenómeno con cierta proporción, se está o no alterando.

Desde el punto de vista fenomenológico y didáctico consideramos que el primer tipo de problemas surge de manera más natural en la práctica. Por ello, en las pruebas piloto, la primera parte del enunciado del primer problema se puede clasificar como problema de este primer tipo.

En realidad, al principio de la investigación no se había considerado la necesidad de restringir el campo de problemas de estadística que queríamos estudiar a los tests de hipótesis. Los problemas del primer tipo se ajustan más a problemas en los que se debe manejar la estimación con base en intervalos de confianza. Por esta razón y a pesar de la importancia que tiene el considerar problemas del primer tipo, decidimos plantear a los estudiantes una segunda prueba piloto en la que sólo se tratara con problemas de la segunda clase.

A continuación presentamos los enunciados de las dos propuestas piloto.

Propuestas piloto

Primera propuesta: “La infidelidad y la iniciativa de separación en las relaciones de pareja”

Uno de los temas que con cierta regularidad se comenta en semanarios y periódicos colombianos es el tema de las relaciones de pareja. Los temas de infidelidad y de separación o divorcio de parejas son “pan de cada día” dentro de una sociedad como la nuestra. Un estudio de opinión sobre los temas anteriores fue recientemente contratado por la revista *7-AL DIA*. Según lo publicado por esta revista, se afirma que la infidelidad ayuda a mantener más estable la relación de pareja y que además —contrario a lo que se podría esperar en una sociedad machista como la nuestra— que las mujeres son más infieles que los hombres. Por otra parte, menciona que la persona que toma la iniciativa de separación en una relación de pareja, es la mujer en un mayor porcentaje que el hombre. Además atribuye a las conductas machistas la causa principal de las separaciones. La revista presentó la siguiente información de carácter técnico.

Características del muestreo
<i>Población de estudio:</i> habitantes casados, separados o en unión libre, de las ciudades de Bogotá, Medellín y Cali.
<i>Tamaño de la muestra:</i> se tomó una muestra de tamaño 784 estratificada según el número de habitantes por ciudad, el sexo, la condición socioeconómica y la edad
<i>Metodología:</i> se elaboró un cuestionario para ser autodiligenciado en forma personal y de manera confidencial
<i>Fecha de aplicación:</i> La información fue recogida entre el 23 de marzo y el 5 de abril de 1995.
<i>Margen máximo de error:</i> 3.5%
<i>Confiabilidad:</i> 95%
<i>Firma encuestadora:</i> OLFELIBE de Colombia

Tabla A.

Parte I. Cuestiones técnicas del muestreo

Para justificar porqué para el margen de error de 3.5%, corresponde un tamaño de muestra de 800, un estadístico explicó que en este muestreo para los diferentes parámetros P de proporciones que se puedan asociar con variables dicótomas del estudio, se calculó un

intervalo de confianza dado en general por $[\hat{P} \pm 0.035]$ con un nivel de confianza del 95%. El valor de \hat{P} depende en particular del parámetro P que se esté estimando, sin embargo, para propósitos de calcular el tamaño de la muestra se puede asumir que el error máximo se comete cuando $p=0.5$.

- 1) Un estudiante afirma que el “margen de error” es precisamente el “error estándar del estimador \hat{P} ” es decir la expresión $\sqrt{\frac{pq}{n}}$. El estudiante está equivocado. Explique porqué y aclare cuál o qué es realmente el margen de error.
- 2) ¿Por qué el estadístico afirma que el error máximo se da cuando $p=0.5$?
- 3) ¿Cómo se obtiene concretamente el valor del margen de error de 0.035?
- 4) Completen la información de la tabla siguiente. En otras palabras, encuentren cuáles deben ser los tamaños de las muestras para los márgenes de error y confiabilidad indicados. **Nota:** Para el cálculo anterior es necesario hacer un supuesto acerca de P . Discútese también a qué supuesto nos estamos refiriendo y porqué es razonable hacerlo.

Error	Confianza	Tamaño
0.1	90%	
0.1	95%	
0.05	90%	
0.05	99%	
0.01	90%	
0.01	95%	
0.01	99%	

- 5) Explique como varían el error y la confianza según cambia el tamaño de la muestra. Para ello debe haber llenado la tabla anterior y observarla con cuidado.
- 6) La revista 7-AL DIA tuvo que considerar tres propuestas de muestreo:
 - a. OLFELIBE de uniandes: fue el presupuesto aceptado y cobraron \$2000 por encuesta trabajando con un margen de error de 0.035 y una confiabilidad de 95%.

- b. NIELSEN de Colombia: cobraban \$500 pesos por encuesta, pero trabajaban con un margen de error de 0.01 y una confiabilidad del 95%.
- c. GALLUP International: cobraban \$1000 pesos por encuesta, trabajando con un margen de error de 0.035 y una confiabilidad de 90%.

¿Si usted fuera el gerente de la revista hubiera aceptado la primera propuesta? ¿por qué?

Parte II: ¿Ha sido usted infiel?

Estudios recientes realizados por una universidad privada mostraron que más del 80% de las mujeres son infieles. La pregunta planteada con la cual se sustentó la investigación fue: “¿Ha sido usted infiel?”. La misma pregunta se envió por correo a profesores de tiempo completo de nuestra Universidad como parte de una encuesta que buscaba explorar la problemática de la relación de pareja al interior de la misma. La encuesta fue respondida por un total de 40 profesores y los resultados de esta encuesta fueron los siguientes:

	Mujeres	Hombres
Respondieron SI	19	16
Respondieron No	1	?

- 1) Con base en los datos anteriores que puede usted concluir con relación al estudio realizado anteriormente.
- 2) ¿Entre qué valores está la proporción real de profesores infieles en nuestra Universidad?
- 3) Investigar qué puede suceder con los valores estimados en el punto anterior si se sabe que la población de profesores de tiempo completo en la Universidad es de 180.

Parte III. ¿Cree usted que la infidelidad ayuda a mantener la relación de pareja?

- 1) La revista Semana de Abril 18 a 25 de 1995 mostró que el 50% de las personas estaban de acuerdo con que la infidelidad ayuda a mantener la relación de pareja. ¿Los datos de la revista 7-AL

DIA son consistentes con los de la revista Semana?

¿La infidelidad ayuda a mantener la relación de pareja?		
	Mujeres	Hombres
Respondieron SI	285	300
Respondieron No	?	50

Según la revista 7-AL DIA entre qué valores porcentuales podemos afirmar que los colombianos no estamos de acuerdo con que la infidelidad ayude a mantener la relación de pareja.

Aplicación

Este problema fue diseñado para trabajar durante tres sesiones de clase de una hora cada sesión. Aunque la idea inicial era trabajar cada una de las partes en cada sesión de clase, sólo se trabajó durante dos sesiones de clase y la tercera parte no se consideró. En la primera sesión se pidió a los estudiantes que resolvieran en grupos de tres estudiantes la parte I del problema y que entregaran, al terminar el tiempo programado, una solución escrita por grupo. La segunda sesión también estaba planeada para trabajar en grupo y así se hizo. El papel del profesor durante la aplicación fue el de apoyar a los grupos que solicitaban ayuda para aclarar dudas. En esta primera prueba no se trabajó con la metodología basada en la corrección de errores.

Evaluación

La aplicación de este cuestionario reflejó la gran dificultad que tenían muchos estudiantes para responder. Algunas de las concepciones erradas que se evidenciaron en las respuestas escritas de los estudiantes y en los diálogos con ellos con relación a la primera parte del enunciado son:

- En la primera pregunta, varios estudiantes asociaban el margen de error con una probabilidad. Por ejemplo, pesaban que como el nivel de confianza era del 95%, entonces había un margen de error del 5%.
- La segunda pregunta no fue entendida por los estudiantes y fue necesaria la intervención del profesor para aclarar el punto.
- Los puntos 3, 4 y 5 fueron resueltos por la mayoría de los estudiantes sin mayores tropiezos. Sin embargo, esto no habría sucedido si el profesor no interviene para aclarar el segundo punto.
- El sexto punto, se hace evidente el descontrol de los estudiantes al querer encontrar una respuesta única a un problema que no tiene respuesta única.

En las respuestas de la segunda parte del problema, encontramos lo siguiente:

- La primera pregunta, que motiva realizar un test de hipótesis, no fue abordada de esta manera por algunos estudiantes que prefirieron utilizar un intervalo de confianza. Por otra parte, dentro del grupo de estudiantes que la respondieron con base en un test de hipótesis, se identificaron varios errores tales como: mal planteamiento de la lateralidad del test e incluso ausencia del planteamiento de hipótesis, interpretación errada de la decisión, etc.
- En la segunda pregunta, los problemas que se presentaron principalmente estaban relacionados con la identificación del tamaño de la muestra. Además, no era claro si se trataba sólo de los profesores hombres o de todos los profesores, hombres y mujeres.
- La tercera pregunta no fue respondida por la mayoría de estudiantes pues no les alcanzó el tiempo para considerarla.

Segunda propuesta

En un artículo que aparece en la revista SEMANA acerca de los índices de violencia en Colombia, se presenta la siguiente información con base en datos suministrados por algunas instituciones como Bienestar Familiar, Policía Nacional, DANE, etc.

	DEPARTAMENTO			
ACCION DE VIOLENCIA	Cund.	Antioq.	Toli.	Valle
Extorsión	63	59	22	68
Terrorismo	21	19	5	27
Secuestro	98	120	27	137
Maltrato Inf.	72	68	32	41

Tabla A. Número de acciones violentas en los deptos. más afectados según los casos registrados en 1994.

	CIUDAD			
ACCION DE VIOLENCIA	Bog.	Med.	Iba.	Cali
Extorsión	8	11	1	13
Terrorismo	2	4	1	3
Secuestro	10	7	6	10
Maltrato Inf.	7	5	3	1

Tabla B. Número de acciones violentas en ciudades capitales con mayor índice de violencia registradas en 1994.

Un grupo de estudiantes de estadística que analizó la proporción “p” de casos de extorsión con respecto a todos los casos de violencia registrados en el departamento de Antioquia, plantearon en un informe de investigación las siguientes hipótesis estadísticas:

$$H_0: p = 26\%$$

$$H_a: p < 26\%$$

De este planteamiento se pudo deducir que la hipótesis de investigación era: *La proporción de casos de extorsión con respecto a los demás actos de violencia en el departamento de Antioquia no supera el 26%*. El mismo grupo de estudiantes planteó con respecto a los actos de secuestro y maltrato infantil las siguientes hipótesis estadísticas:

Caso A	Caso B
Secuestro en Ibagué	Maltrato infantil en el Tolima
$H_0: p = 35\%$	$H_0: p = 20\%$
$H_a: p > 35\%$	$H_a: p \neq 20\%$

Para cada caso:

- 1) Determine cuál fue la hipótesis de investigación que los estudiantes debieron haber planteado.
- 2) Defina formalmente cuál es el parámetro de estudio.
- 3) Realice el test de hipótesis planteado. Utilice en nivel de significación del 1%.
- 4) Encuentre un nivel de significación tal que la interpretación del resultado del problema cambie de sentido. Es decir, determine un nivel de significación tal que si no rechazó H_0 , ahora si rechace H_0 o viceversa, si rechazó H_0 ahora no se rechace H_0 .
- 5) ¿Se justifica hacer el cambio del nivel de significación que se sugiere en el punto anterior? Explique.

Aplicación

Este problema fue diseñado para trabajar durante dos sesiones de una hora cada una. En la primera sesión se pidió a los estudiantes que resolvieran de manera individual el problema enunciado y lo entregaran por escrito. Previo a la segunda sesión estaba planeado identificar lugares donde había errores de los estudiantes, pero sin mencionar en que consistía el error.

La segunda sesión estaba planeada para trabajar en grupo. Los estudiantes debían revisar en grupos de tres estudiantes las respuestas que habían presentado de manera individual y producir un nuevo informe escrito por grupo. Fue en esta segunda prueba piloto cuando se empezaron a realizar ensayos para implementar estrategias metodológicas basadas en la corrección de errores.

Evaluación

El problema planteado es demasiado largo. Los estudiantes sólo pudieron resolver uno de los dos casos en la sesión de trabajo individual. Por ello fue

necesario descartar el que los estudiantes presentaran la solución del segundo caso. En la segunda sesión se trabajó en la corrección del caso que se había presentado de manera individual. Se identificaron errores tales como:

- Falta de claridad para identificar el parámetro de estudio.
- Utilizar la aproximación normal de la binomial sin revisar los supuestos de aproximación.
- Se utiliza el término nivel de confianza en lugar de nivel de significación.

Propuesta definitiva

Los resultados obtenidos en los ensayos piloto obligaron a replantear de nuevo no sólo la formulación de las preguntas sino la metodología de aplicación. El cuestionario definitivo se dividió en dos partes, una para aplicar de manera individual (una hora) y la otra para trabajar en grupo (una hora). La segunda parte esencialmente es igual que la primera. En realidad en la segunda parte se cambió la lateralidad de las hipótesis del test y los tamaños muestra llevan a que se deba utilizar el modelo binomial. Los estudiantes podrían utilizar las calculadoras gráficas y tenían un entrenamiento previo en su manejo. Los enunciados de esta propuesta definitiva se presentan a continuación.

Primer cuestionario (parte individual)

En un artículo que aparece en la revista SEMANA se habla acerca de los índices de violencia en Colombia. Dicho artículo presenta la siguiente información:

ACCIONES DE VOLENCIA	DEPARTAMENTOS	
	Antioquia	Tolima
Extorsión	59	22
Terrorismo	19	5
Secuestro	120	27
Maltrato Infan.	68	32

Tabla A. Número de acciones violentas en los deptos. más afectados según los casos registrados en 1994.

Algunos investigadores que trabajan en el tema de la violencia afirman que:

Los casos de extorsión con respecto a todos los casos de violencia registrados en el departamento de Antioquia, no superan el 26%.

Con base en los datos suministrados por la revista semana, analice la afirmación anterior. Para ello, resuelva los siguientes puntos:

- 1) Formule las hipótesis nula y alternativa correspondientes
- 2) Encuentre el p-valor y represéntelo gráficamente.
- 3) Determine un nivel de significación tal que se rechace H_0
- 4) Determine un nivel de significación tal que no se rechace H_0 .
- 5) ¿Qué decisión tomaría con respecto a la afirmación? Explique claramente su respuesta

El Gobierno Nacional ha destinado un presupuesto de 500 millones de pesos para el departamento de Antioquia. El presupuesto sólo se puede invertir en educación o en seguridad, pero no en ambos. La inversión se puede hacer en equipos y entrenamiento de personal para reforzar al Grupo Gaula (grupo policivo especializado en antiterrorismo) o en colegios públicos, para materiales y capacitación de profesores que favorezcan el desarrollo de P.E.I's (proyectos educativos institucionales) de mayor calidad. Suponga que la decisión sobre en qué área (educación o seguridad) se debe invertir el presupuesto, depende de si han aumentado o no los niveles de violencia.

- 6) Analice, según el contexto, cuál es la incidencia de los errores de tipo I y tipo II.

Segundo cuestionario (parte de trabajo en grupo)

En un artículo que aparece en la revista SEMANA se habla acerca de los índices de violencia en Colombia. Dicho artículo presenta la siguiente información:

ACCIONES DE VIOLENCIA	CIUDADES	
	Bogotá	Ibagué
Extorsión	8	1
Terrorismo	2	1
Secuestro	10	6
Maltrato Infan.	7	3

Tabla B. Número de acciones violentas en ciudades capitales con mayor índice de violencia registradas en 1994.

Algunos investigadores que trabajan en el tema de la violencia afirman que:

Los casos de secuestros en la ciudad de Ibagué son mayores que el 35%.

Con base en los datos suministrados por la revista semana, analice la afirmación anterior. Para ello, resuelva los siguientes puntos:

- 1) Formule las hipótesis nula y alternativa correspondientes
- 2) Encuentre el p-valor y represéntelo gráficamente.
- 3) Determine un nivel de significación tal que se rechace H_0
- 4) Determine un nivel de significación tal que no se rechace H_0 .
- 5) ¿Qué decisión tomaría con respecto a la afirmación? Explique claramente su respuesta

El Gobierno Nacional ha destinado un presupuesto de 200 millones de pesos para el departamento de Tolima. El presupuesto sólo se puede invertir en educación o en seguridad, pero no en ambos. La inversión se puede hacer en equipos y entrenamiento de personal para reforzar al Grupo Gaula (grupo policivo especializado en antiterrorismo) o en colegios públicos, para materiales y capacitación de profesores que favorezcan el desarrollo de P.E.I's (proyectos educativos institucionales) de mayor calidad. Suponga que la decisión sobre en qué área (educación o seguridad) se debe invertir el presupuesto, depende de si han aumentado o no los niveles de violencia.

- 6) Analice, según el contexto, cuál es la incidencia de los errores de tipo 1 y tipo 2

Conceptos y procedimientos que se ponen en juego

En la Tabla N° 7 se presentan los conceptos y los procedimientos que supuestos que se debían utilizar para resolver el problema planteado. Se ha señalado al frente de cada uno de los procedimientos contemplados la pregunta del problema que motiva su utilización. En la misma Tabla N° 7 también se muestra la unidad conceptual que se asocia a cada conjunto de conceptos y procedimientos y la fase de resolución de problemas en donde supuestamente se ponen en juego.

Tabla N° 7. Conceptos, y procedimientos que se ponen en juego, en el cuestionario

Conceptos, unidad conceptual y fase involucrada	Procedimientos asociados	Item
Variable Unidad A Fase I	A1. Identificar la variable de estudio como binomial.	1
	A2. Identificar la variable continua.	1
Muestra Unidad A Fase I	A3. Distinguir la población de la muestra.	2
Nivel de significación Unidad A Fase I	A4. Determinar un nivel de significación de modo que se rechace o no H_0 .	3 y 4
Parámetros Unidad B Fase I	B1. Identificar como proporción al parámetro de estudio.	1
Hipótesis de investigación Unidad B Fase I	B2. Identificar la hipótesis de investigación del problema.	1
Hipótesis estadísticas Unidad B Fase I	B3. Formular las hipótesis refiriéndose a la proporción poblacional.	1
Lógica del contraste Unidad B Fase I	B4. Establecer el contraste de hipótesis como un problema de decisión entre dos hipótesis estadísticas.	1
Estadística de prueba Unidad C Fase II	C1. Calcular el valor de la proporción muestral estandarizarlo y ubicarlo en un gráfico.	2
Distribución del estadístico de prueba Unidad D Fase II	D1. Identificar la distribución asociada al estadístico de prueba.	2
	D2. Determinar el α de un valor de la estadística de prueba y graficarlo.	2
	D3. Representar gráficamente las regiones de rechazo y no rechazo de H_0 .	3, 4
Nivel de significación Unidad E Fase III	E1. Comparar el α con un p-valor para tomar una decisión.	3, 4 y 5
Error de tipo 1 Unidad E Fase III	E2. Evaluar la incidencia del error de Tipo 1.	6
Error de tipo 2 Unidad E Fase III	E3. Evaluar la incidencia del error de Tipo 2.	6

Por otra parte, en la Tabla N° 8 se presentan al frente de las interrelaciones que se ponen en juego los ítems o las partes del enunciado donde de manera implícita o explícita se debe tener presente dicha interrelación. En la última columna se hace un breve comentario acerca de la interrelación en relación con el ítem identificado.

Tabla N° 8. Interrelaciones que se deben tener presentes en el cuestionario

N°	Interrelaciones	Puesta en juego en	Comentario
1	En la formulación del contraste de hipótesis se está ante un problema de decisión entre dos hipótesis de las cuales no sabemos cuál es verdadera.	Enunciado	Aparece de manera implícita en el enunciado del problema: “Algunos investigadores que trabajan en el tema de la violencia afirman que los casos de extorsión con respecto a todos los otros casos de violencia registrados en el departamento de Antioquia no superan el 26%”
2	Hay que diferenciar entre hipótesis nula e hipótesis alternativa. Cada una de ellas juega un papel diferente en el contraste.	1	El punto pide establecer esta diferencia de manera explícita
3	Hay que diferenciar entre hipótesis unilaterales y bilaterales. El procedimiento de contraste se ve afectado por esta distinción.	1, 3, 4	Hay que establecer esta diferencia de manera explícita
4	Existe la posibilidad de cometer un error de tipo I: rechazar una hipótesis nula verdadera.	6	Hay que establecer esta diferencia de manera explícita
5	El riesgo del error de tipo I viene medido por el nivel de significación.	3	Aparece de manera implícita al pedir un valor para el nivel de significación que cumpla con la condición pedida se evalúa el riesgo del error.
6	Existe la posibilidad de cometer un error de tipo II: aceptar una hipótesis nula falsa.	6	La decisión lleva a que si no se rechaza se está en posibilidad de cometer error tipo II y se debe evaluar la consecuencia.

Tabla N° 8. Interrelaciones que se deben tener presentes en el cuestionario

N°	Interrelaciones	Puesta en juego en	Comentario
7	Las hipótesis paramétricas se refieren a valores de los parámetros de las poblaciones	1	El punto pide establecer esta diferencia de manera explícita
10	El estadístico de prueba es una estandarización del estimador	2	Para determinar el p-valor es necesario evaluar el estadístico de prueba.
12	El estadístico es una función de los valores de la muestra, es una variable aleatoria	2	Al utilizar la aproximación de la distribución binomial a la distribución normal se requiere de la proporción muestral para determinar el estadístico de prueba.
14	La distribución del estadístico depende del valor asociado a parámetros de la población.	1	Se debe considerar el tamaño de la muestra para determinar el tipo de distribución asociada.
15	El nivel de significación es el área determinada por el valor crítico en la función de densidad de la distribución del estadístico en el muestreo, bajo la hipótesis nula.	3 y 4	El punto pide establecer esta diferencia de manera explícita
16	Un resultado estadísticamente significativo es un resultado cuya probabilidad de ocurrencia, en caso de ser cierta la hipótesis nula es menor que el nivel de significación.	5 y 6	Para tomar una decisión se deben comparar el p-valor y el nivel de significación.
17	La significación estadística no implica necesariamente la significación (relevancia) desde el punto de vista práctico.	5	Al tomar una decisión se debe tener presente la incidencia del tipo de error para el problema.
21	La distribución de un estadístico suele depender del tamaño de la muestra.	2, 3 y 4	El punto pide establecer esta diferencia de manera explícita

Tabla N° 8. Interrelaciones que se deben tener presentes en el cuestionario

N°	Interrelaciones	Puesta en juego en	Comentario
25	El nivel de significación, junto con las hipótesis nula y alternativa y el estadístico de prueba determinan las regiones crítica y de aceptación en un contraste.	1, 3 y 4	El punto pide establecer esta diferencia de manera explícita
26	El p-valor del resultado de un estadístico de prueba depende de los datos de la muestra, de la formulación de las hipótesis y de la distribución del estadístico	1, 2	El punto pide establecer esta diferencia de manera explícita
27	El nivel de significación proporciona un criterio de decisión probabilístico, no un criterio matemático certero e incuestionable	3 y 4	Cuando se pide determinar un nivel de significación que cumpla con la condición dada sólo se espera que se rechace o no una hipótesis.

Una vez establecidos los conceptos, los procedimientos y las interrelaciones que se ponen en juego en la situación problemática propuesta, pasamos a considerar la parte donde se quiere describir las posibles dificultades asociadas a los errores identificados y algunas de las formas prevalentes de comprensión.

Errores a priori

A continuación describimos los errores que pensamos que sucederían antes de proceder a la aplicación de la propuesta definitiva. Para organizar la descripción utilizamos las cinco unidades conceptuales que se describieron en el análisis de contenido.

Relacionados con la unidad llamada ‘Datos’:

- No diferencia la población de la muestra.
- No distingue la proporción de la población de la proporción de la muestra.

Relacionados con la unidad llamada ‘Formulación de hipótesis’:

- No reconoce el parámetro de estudio.
- No reconoce la lateralidad del test.
- Plantea las hipótesis refiriéndose al estimador y no al parámetro.

- Omite el signo “=” en la hipótesis nula y /o lo pone en la hipótesis alternativa.

Relacionados con la unidad llamada ‘Herramientas’:

- Identifica mal el estimador: toma un tamaño de la muestra equivocado.
- No calcula correctamente el valor del estadístico de prueba.

Relacionados con la unidad llamada ‘Distribuciones’:

- No identifica correctamente el nivel de significación.
- Representa de manera errada las regiones de rechazo y de no rechazo de “Ho.”

Relacionados con la unidad llamada ‘Análisis e interpretación’:

- Compara probabilidades con valores z para tomar una decisión.
- Interpreta mal el error de tipo I o el de tipo II.

Manejo de representaciones externas

A continuación hacemos una revisión de algunas de las traducciones en relación con la forma como se pone en juego el concepto de nivel de significación en la propuesta definitiva. Como veremos, la calculadora gráfica puede jugar un papel importante en la resolución de problema, al facilitar las representaciones tabulares y gráficas de las distribuciones binomial y normal.

1. Traducción de tablas a situaciones (descripciones verbales). En este sentido los puntos 3 y 4 del cuestionario piden explícitamente determinar un nivel de significación para rechazar (punto 3) y no rechazar (punto 4) la hipótesis nula. Se debe buscar en la tabla estos niveles y escribir la explicación en el papel. Aquí la calculadora gráfica puede jugar un papel intermedio. Si el estudiante luego de obtener el valor de la estadística de prueba, lo introduce en la calculadora para producir una representación gráfica, tendríamos en juego una representación más y un procedimiento de interpretación apoyando el razonamiento.

2. Traducción de situaciones (descripciones verbales) a tablas. En esta dirección es como usualmente se plantea la búsqueda no del nivel de significación, sino del valor crítico correspondiente al nivel de significación dado. Este tipo de traducción no se plantea directamente en ninguno de los puntos del problema propuesto.

3. *Traducción de gráficas a situaciones (descripciones verbales)*. Como no ocurre normalmente estamos esperando que ocurra por intermedio de la calculadora gráfica como ya se comentó en el primer caso.

4. *Traducción de situaciones (descripciones verbales) a gráfica*. Cuando en el punto 2 se pide que se represente gráficamente el p-valor, se esperaría que el estudiante lo tuviera como referencia para compararlo con el nivel de significación.

5. *Traducción de situaciones (descripciones verbales) a fórmulas*. No se dan en este problema

6. *Traducción de tablas a gráficas*. En este caso la calculadora gráfica proporciona un cambio en los medios de representación. En lugar de las tradicionales tablas de la distribución binomial impresas en los textos, puede obtener una representación tabular en la calculadora que luego puede trasladar a un gráfico en el papel. Sin embargo, el estudiante también puede obtener una representación gráfica en la calculadora. En este caso es posible que el estudiante haga una representación gráfica sin pasar por la representación tabular, pero no tendrá los valores de las probabilidades, que son esenciales para encontrar un p-valor o para encontrar valores críticos asociados a un nivel de significación.

7. *Traducción de gráficas a tablas*. La representación gráfica en la calculadora gráfica le aporta al estudiante una manera de verificar si está realizando bien la lectura de una tabla de distribuciones.

ANÁLISIS DE ENSEÑANZA

Dentro del currículo del curso este tema se desarrolló en la tercera semana de clase, luego de ver el tema de inferencia acerca de la proporción de una población. El profesor usualmente había desarrollado este tema luego de ver el tema de inferencia acerca de la diferencia de medias. Sin embargo, para este semestre y debido al cambio de texto que se propuso se decidió desarrollar en primer lugar este tema. Desafortunadamente, el texto nuevo traía desarrollado el tema de test de hipótesis para la proporción únicamente con la aplicación de la distribución normal. Por ello el profesor, sin dejar totalmente de lado el texto oficial del curso, decidió utilizar la parte del texto [3] que contenía este desarrollo y dar más énfasis al uso de éste para el estudio de este tema.

Comentarios acerca del profesor y el texto utilizado

El texto [1] hace una presentación general en la cual sólo se menciona el uso de la aproximación de distribuciones binomiales por medio de la distri-

bución normal. En [1] sólo se explica el procedimiento para realizar un test de hipótesis con base en la aproximación normal de la binomial y se usan las aproximaciones clásica y al p-valor. Por contraste en el texto [3], se hace una caracterización de las distribuciones binomiales así como la ejemplificación de tests de hipótesis realizadas tanto con distribuciones binomiales como con la aproximación normal de la distribución binomial. Sin embargo, en [3] no se explica la aproximación al p-valor.

La mayoría de estudiantes se inclinaron más por la lectura del texto [3]. Lo cual se puede explicar por el énfasis que hizo el profesor en las tests de hipótesis con muestras pequeñas donde había que recurrir directamente a la distribución binomial. Además, en el texto [1] no hay problemas de tests de hipótesis que se deban resolver con base en la distribución binomial ni ejemplos resueltos a manera de ilustración.

Para resumir en pocas palabras, se podría decir que el texto [1] es demasiado corto y escueto en la presentación de este tema, mientras que el texto [4], peca por lo contrario, es demasiado largo y a pesar de esto no presenta la aproximación al p-valor.

Características de la aplicación de la situación problemática

Algunas de las características que están presentes en la aplicación de la situación problemática son las siguientes:

- La forma de trabajo que se pide aplicar en la primera sesión es individual y en la segunda sesión es en grupo.
- La aplicación individual motiva el uso de la distribución normal y la aplicación en grupo el de la distribución binomial. La aplicación en grupo no se enfoca de manera primordial en la discusión de los errores, pero advierte de ellos para la realización del cuestionario de trabajo en grupo.
- Ha habido un entrenamiento previo de tres horas de clase para manejar la calculadora gráfica.
- La calculadora gráfica puede ser utilizada para encontrar y representar de manera gráfica el p-valor.
- El acceso a la calculadora gráfica permite a los estudiantes evaluar tanto los resultados aritméticos como los gráficos y tabulares.
- El profesor no debe intervenir en la aplicación individual. Debe identificar los errores que se presentan en los papeles que le entregan los estudiantes al final de la aplicación individual y producir un reporte de dichos errores para la parte de trabajo en grupo que deberá entregar a los estudiantes.

- El reporte de errores tiene un doble propósito. Por un lado, el de intentar que los estudiantes reconozcan, en lo posible, las equivocaciones que cometieron en la parte individual y no las repliquen en el trabajo escrito que presenten en grupo. Y por otro lado, hacerle un seguimiento al tipo de errores que comete cada estudiante en particular.

EVALUACIÓN

Errores, dificultades y obstáculos

Errores a posteriori

En general, ocurrieron los errores esperados a priori que se comentaron en la sección anterior. Sin embargo, también se identificaron nuevos errores que no se habían previsto. A continuación relacionamos dichos errores.

Relacionados con la unidad llamada ‘Datos’ se encontró que algunos estudiantes:

- Confunden el tamaño de la muestra con el número de éxitos o de fracasos.

Relacionados con la unidad llamada ‘Formulación de hipótesis’ se encontró que algunos estudiantes:

- Confunden la notación “ p ” utilizada para representar el parámetro de la proporción de una población, con la notación “ μ ” utilizada para representar el parámetro de la media de una población.

Relacionados con la unidad llamada ‘Distribuciones’ se encontró que algunos estudiantes:

- No calculaban correctamente el p -valor.
- Hacían representaciones gráficas incompletas o incorrecta del p -valor.
- No encontraban correctamente el valor numérico del “ α ” correspondiente a un número de éxitos específico.

Relacionados con la unidad llamada ‘Análisis e interpretación’ se encontró que algunos estudiantes:

- No podían determinar un nivel de significación para rechazar o no rechazar la hipótesis nula.

Reporte de la lista de errores

Luego de que los estudiantes entregaron las soluciones escritas al problema planteado para trabajar de manera individual, se revisaron las soluciones y se elaboró un reporte para presentarles a los estudiantes, al día siguiente, la descripción de los errores que se habían identificado. A continuación presentamos dicho reporte, tal como se les presentó a los estudiantes:

Errores de los estudiantes en la situación problemática N° 1:

- E1: Confunde el parámetro con el estimador o con otro parámetro
- E2: No están bien planteadas las hipótesis estadísticas.
- E3: Hay una representación gráfica incompleta o incorrecta del nivel de significación.
- E4: No presenta el p-valor que corresponde al valor de la estadística de prueba
- E5: Está mal calculado el valor del estimador de la proporción de la población.
- E6: No se calculó, o se calculó de manera incorrecta, el valor de la estadística de prueba.
- E7: Hay una determinación incorrecta del p-valor.
- E8: No se están comparando de manera correcta los resultados obtenidos con los valores críticos.
- E9: No se determina correctamente el nivel de significación que es necesario para rechazar, o no rechazar, la hipótesis nula.

Para resumir esta información se elaboró una tabla donde está contenida la información, por alumno, de los errores que les fueron señalados a cada uno de ellos en su escrito. En esta tabla están resaltadas en negrilla las tres estudiantes que fueron observadas en detalle en la interacción del trabajo en grupo. A continuación presentamos la tabla mencionada.

Tabla N° 9.

Estudiante	E1	E2	E3	E4	E5	E6	E7	E8	E9	Total de errores
Faihan	•	•		•					•	4
Hugo		•					•		•	3
Andrés			•	•						2

Tabla N° 9.

Estudiante	E1	E2	E3	E4	E5	E6	E7	E8	E9	Total de errores
Juan		•	•			•				3
Ximena		•			•				•	3
Carlos		•	•		•	•				4
Norma		•	•			•	•			4
Gelber	•	•		•					•	4
Angela		•	•				•	•		4
Martha		•			•	•	•	•	•	6
Natalia		•	•		•	•				4
Carolina		•	•		•	•				4
Frecuencia de errores	2	11	7	3	5	6	4	2	5	45
Porcentaje	16.6%	91.6%	58.3%	25%	41.6%	50%	33.3%	16.6%	41.6%	

Con referencia a la tabla anterior se encuentra que el error más frecuente está relacionado con el planteamiento de las hipótesis estadísticas. En algunos casos, como en el del problema que se planteó, al estudiante le es difícil establecer la hipótesis nula y la hipótesis alternativa cuando no son explícitas en el enunciado del problema; así mismo, se pueden encontrar dificultades en la comprensión de la igualdad en la hipótesis nula. La mayoría de los estudiantes simplemente aprendieron que el igual debe ir en la nula, sin saber explicar porqué.

En cuanto a la representación gráfica del p-valor surgieron dificultades con el uso de las tablas. En general, prefirieron el uso de la calculadora gráfica, para determinar la distribución. Sin embargo, pocos estudiantes utilizaron la calculadora para representar en forma gráfica la distribución. Por otra parte, la dificultad para discriminar entre p-valor y nivel de significación se hizo evidente en este problema, lo mismo que la preferencia por utilizar la aproximación clásica sobre la aproximación al p-valor.

Con base en los resultados obtenidos, también se evidenció la falta de comprensión de algunas de las preguntas del enunciado.

Dificultades y obstáculos

Se supone que los estudiantes que se encuentran en un segundo curso de estadística, tienen la capacidad suficiente para aplicar su conocimiento en la producción de respuestas que normativamente se consideran correctas; sin embargo, existe la posibilidad que algunos los de errores relacionados principalmente con la formulación de las hipótesis, con el manejo de las distribuciones y con el análisis e interpretación de datos puedan tener su

origen en una forma prevalente de pensamiento que no necesariamente es correcta.

Hemos ubicado en el gráfico que representa la estructura conceptual de contenidos del test de hipótesis, los tipos de error que se identificaron en los estudiantes. Esta ubicación de los errores sugiere cuáles son las conexiones entre las unidades conceptuales que están más débiles. Por ejemplo, la conexión entre la unidad Distribuciones y la unidad Análisis parece que es de las más débiles, allí se relacionan los errores 4, 8 y 9. Por otra parte, el error 2, que no se ha conectando con otras unidades, pensamos que en realidad se debe a la falta de comprensión de la lógica que subyace a los test de hipótesis.

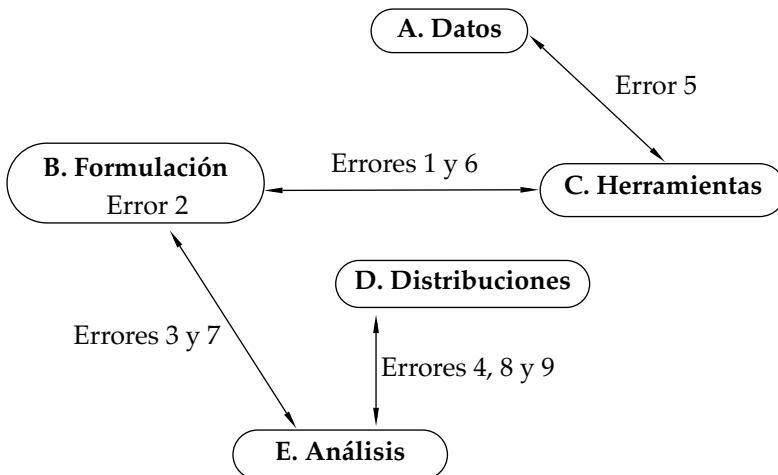


Figura 4.

Por otra parte, el error 8 es uno de los que posteriormente ocurrió con más persistencia. El hecho de que se comparen valores de probabilidades con valores percentiles (por ejemplo, valores de “z” de la distribución normal) pensamos que se puede deber a que probablemente sólo existe en los estudiantes una noción de comparación de orden restringida a lo puramente lineal. De este modo, cuando se ponen en juego en el problema los representantes numéricos que corresponden a probabilidades (p-valor o nivel de significación) se representan en forma lineal. Por eso, para ellos puede ser completamente válido que la comparación 0.05 (valor de un nivel de significación) < 1.42 (valor de un percentil) sea válida independientemente de que 0.05 corresponda a una probabilidad y 1.42 a un valor “z”.

Ahora vamos a considerar ejemplos concretos de errores de los estudiantes y a comentar algunas de las dificultades que supuestamente podrían incidir en el mismo.

Error 1. Confusión del parámetro con el estimador en la notación y/o en los cálculos. Gelber escribió el planteamiento de las hipótesis estadísticas así:

Ho: $\hat{P} \geq 0.26$ y Ha: $\hat{P} < 0.26$; Y Faihan escribió: Ho: $P \geq 0.22$ y Ha: $P < 0.22$.

Error 2. Planteamiento incorrecto de las hipótesis estadísticas bien sea en forma verbal o en forma simbólica: omisión o incorrecta ubicación del “=” o planteamiento inverso de la lateralidad. La mayoría de estudiantes planteó las hipótesis así: Ho: $P > 0.26$ y Ha: $P < 0.22$. Norma colocó el “=” en la hipótesis alternativa.

Error 3. Representación gráfica incompleta o incorrecta del nivel de significación. Hugo hizo un dibujo como el que se muestra a la izquierda y Martha hizo un dibujo como el que se muestra a la izquierda.

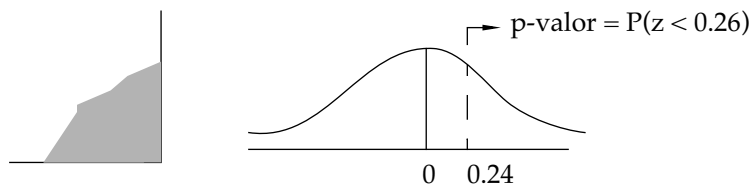


Figura 5.

Error 5. Cálculo incorrecto del estimador de la proporción de la población: no identifica el valor del tamaño de la muestra o lo confunde con otro valor. En el escrito individual, Natalia tomó el valor del tamaño de la muestra como $n = 266 + 86 = 352$ en lugar de tomarlo como $n = 266$.

Error 6. Cálculo incorrecto del valor de la estadística de prueba o ausencia del cálculo. Por ejemplo Juan, al evaluar el estadístico de prueba, utilizó el valor del estimador en lugar del valor del parámetro dado por la hipótesis nula. En general, cuando se trata de evaluar el estadístico de prueba los errores más frecuentes se encuentran a nivel de cálculos. En muchos casos esta situación se debe a que identifican mal el estimador.

Error 7. Determinación incorrecta del valor del p-valor. Por ejemplo, Hu - go encontró el p-valor así: $p\text{-valor} = 1 - 0.26 = 0.74$.

Error 8. Comparación incorrecta de valores críticos o del p-valor con el nivel de significación. Por ejemplo, Martha Compara valores de z con p-valores. escribió: “si utilizo un nivel de significación de 0.05, el p-valor = 0.24 es muy pequeño con respecto a 1.65, entonces se rechazaría la hipótesis nula”. En varios casos, en la representación gráfica de las regiones de

rechazo y no rechazo de la hipótesis nula, se evidencia la falta de claridad con respecto a la ubicación que se debe dar al nivel de significación, al p-valor, al valor normal correspondiente al nivel de significación y al valor de la estadística de prueba. Pero esta situación se hace aún más evidente en los casos de estudiantes que comparan valores críticos con probabilidades para tomar una decisión con respecto a la hipótesis nula.

9. SITUACIÓN PROBLEMÁTICA N° 2: INFERENCIA ACERCA DE LA DIFERENCIA DE MEDIAS DE DOS POBLACIONES

INTRODUCCIÓN

En esta parte se presenta el trabajo que se realizó para el diseño, desarrollo y evaluación de la segunda situación problemática. En esta ocasión se trabajó el tema de test de hipótesis sobre la comparación de la media de dos poblaciones, con base en dos muestras independientes. Al igual que en el diseño de la situación anterior se utiliza un esquema metodológico que se basa en la explicación y la corrección de errores por parte del estudiante.

Para la concepción y concreción del diseño se hace un análisis didáctico que contempla tres dimensiones: por un lado, se presenta un análisis del contenido estadístico del tema siguiendo el esquema planteado en las partes (capítulos) 2 y 3. Luego se presenta un análisis del aprendizaje en el que concretamente se elige un tipo de situación problemática y se mira en ella los conceptos y procedimientos que se ponen en juego, se conjetura acerca de los posibles errores que podrían cometer los estudiantes al abordarlo y se miran algunos aspectos del lenguaje estadístico utilizado con base en el análisis de representaciones externas. Luego, se considera la dimensión de la enseñanza atendiendo principalmente al análisis del texto que leyeron los estudiantes antes de enfrentarse al problema propuesto e intentando establecer una relación entre este análisis y los errores conjeturados. Al final, se presenta la evaluación de la aplicación de la situación problemática. Esta evaluación se hace principalmente con base en el análisis de los errores.

La forma de aplicación de la situación problemática se puede resumir en tres pasos. Primero, los estudiantes deben resolver la situación de manera individual y entregarla por escrito. Segundo, los profesores la revisan e identifican y señalan los errores de los estudiantes y luego planean la sesión de trabajo en grupo de los estudiantes. Tercero, los estudiantes se reúnen de nuevo en la clase, trabajan en grupo y su tarea consiste en la explicación y en la corrección de los errores que se les ha señalado. Además, se mira con más detalle la forma como uno de los grupos de tres estudiantes explica y corrige los errores.

ANÁLISIS DEL CONTENIDO

A continuación presentamos algunos de los elementos conceptuales del conocimiento estadístico relacionado con este tema.

A. Datos

Variables

El concepto de variable asociado con estos problemas de test de hipótesis se puede ver desde varios puntos de vista. Si miramos las variables que intervienen desde el punto de vista metodológico, se debe considerar por un lado, la presencia de una variable independiente dicótoma que identifica o discrimina las dos poblaciones específicamente consideradas, y por el otro, la presencia de una variable dependiente a partir de la cual se definen los parámetros de cada población. La variable dependiente es de tipo cuantitativo y continuo, aunque hay en la práctica algunas ocasiones donde también se manejan variables discretas. Si se miran las variables desde el punto de vista del tipo de medición, la variable dependiente se ubica dentro de la categoría nominal y la dependiente dentro de las escalas de intervalo o de razón.

Población

Deben existir dos poblaciones las cuales se pueden distinguir según la variable independiente. Con el test de hipótesis se busca determinar si estas poblaciones también se pueden o no distinguir cuando se infiere acerca del valor de la media de la variable dependiente en cada población.

Muestras

Se manejan dos muestras independientes. Para establecer si las muestras son o no independientes se debe verificar que las mediciones realmente se han hecho sobre sujetos o entes diferentes. Las muestras deben ser obtenidas de manera aleatoria y el tamaño de las muestras debería ser proporcional al tamaño de las respectivas poblaciones.

Nivel de significación

Visto como dato, puede o no darse a priori en los datos del enunciado de un problema. Su relevancia como tal está ligada al contexto del problema que se plantee. Sin embargo, a diferencia de los problemas donde se trabaja con distribuciones discretas, siempre se puede encontrar para el nivel de significación que sea dado, el valor crítico correspondiente.

Medidas de resumen estadístico

Las medidas de resumen que son relevantes en este tipo de inferencia son la media muestral y la varianza muestral de cada una de las dos muestras. Cuando se mira o se propone el enunciado de un problema, estas medidas ya pueden estar calculadas o puede suceder que sea necesario calcularlas a partir de los datos propios de las muestras.

B. Formulación

Parámetros

La diferencia de las medias de dos poblaciones es el parámetro sobre el que se centra la inferencia. En términos de la variables dependiente e independiente que se asocian a un problema de diferencia de medias, cada uno de los parámetros de esta diferencia se puede definir como la media aritmética de la variable dependiente de los datos de cada una de las respectivas poblaciones.

En este caso es importante resaltar que cada una de las medias se define sobre una población de estudio diferente y no, como usualmente sucede en la comparación de muestras relacionadas, sobre dos poblaciones de datos distintas asociadas a una misma población de estudio.

Hipótesis de investigación

Usualmente la hipótesis de investigación sugiere que los parámetros no tienen el mismo valor, es decir, que el valor de cada parámetro no es el mismo. Incluso, algunas veces se dice explícitamente que el valor de uno de los parámetros es mayor que el del otro. No es tan usual plantear el que la diferencia de los parámetros sea un valor fijo diferente de cero, aunque la teoría también permite abordar este tipo de problemas.

Hipótesis estadísticas

La costumbre más extendida desde el punto de vista de la notación es utilizar μ_1 y μ_2 para denotar el valor hipotético de cada parámetro poblacional. El enfrentamiento de estas dos hipótesis aparece, por ejemplo, de la siguiente manera: $H_0: \mu_1 - \mu_2 \geq 0$ vs. $H_a: \mu_1 - \mu_2 < 0$.

A diferencia de los problemas de una muestra, donde generalmente se especifica el valor concreto de una posible media de la población, en este caso no importa saber los valores concretos de las medias. Sólo interesa establecer si son iguales o diferentes o si una media es mayor que la otra.

Lógica del contraste

Dado que usualmente la hipótesis de investigación sugiere que los parámetros no tienen el mismo valor, la hipótesis nula debe, en general, asumir el valor cero para la diferencia de las medias poblacionales. Los casos como el que se mencionó anteriormente donde en el contraste de hipótesis se tenga una diferencia diferente de cero en la hipótesis nula, estarían aparentemente, en contradicción con el sentido de "nulidad" o de la no diferencia en una hipótesis nula.

C. Herramientas

Estimadores

El interés se centra no tanto en la estimación de cada uno de los parámetros μ_1 y μ_2 , sino en la estimación de la diferencia entre μ_1 y μ_2 . La estimación se hace con base en la diferencia de promedios muestrales dada por $\bar{X}_1 - \bar{X}_2$.

Varianza del estimador $\bar{X}_1 - \bar{X}_2$

En lo que se refiere a la variabilidad, cuando no se conocen las varianzas de las dos poblaciones y se quiere analizar la dispersión de los datos con base en las muestras, es necesario estimar las desviaciones estándar poblacionales σ_1 y σ_2 con S_1 y S_2 respectivamente. Bajo el supuesto de homocedasticidad en el que se asume que $\sigma_1 = \sigma_2 = \sigma$, se tendrían dos estimaciones S_1 y S_2 de σ posiblemente diferentes. En este caso se calcula una varianza ponderada $(S_p)^2$, con base en las desviaciones S_1 y S_2 .

Por lo tanto el error estándar que se asocia al estimador presenta diferentes posibilidades dependiendo de la forma como vengan dados los datos y del supuesto de homocedasticidad. Por ejemplo, cuando se supone que la varianzas de las poblaciones son iguales y se estima la varianza ponderada, el error estándar se debe calcular así:

$$\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \cdot \left(\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\right)}$$

Sin embargo, si se supone que se conocen las varianzas de las poblaciones el error estándar se debe calcular así:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Finalmente, si es necesario utilizar las varianzas estimadas de cada población, pero no parece que sea razonable asumir el supuesto de homocedasticidad, el error estándar se debe calcular así:

$$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Estas tres variantes tornan más complejo el manejo del error estándar y, como veremos en el análisis de la enseñanza, se deben considerar con cuidado las explicaciones y decisiones de orden didáctico en torno de ellas.

Estadísticos para tests de hipótesis

Con respecto a la estadística de prueba y al igual que en muchos otros tests paramétricos su cociente de estandarización viene dado por la razón entre la diferencia del estimador muestral de las medias poblacionales y su respectivo error estándar, según los supuestos que se hayan considerado.

D. Distribuciones

La distribución de las poblaciones de las cuales se extraen las muestras, debería ser normal si se requiere hacer inferencia con base en muestras pequeñas. Cuando se está estimando la varianza de la población, la teoría estadística establece que es necesario este supuesto para que la distribución límite del promedio muestral siga la distribución t-student. Esta distribución juega un papel central en este tipo de problemas.

La diferencia de medias muestrales sigue, dependiendo de los supuestos asociados a los teoremas de distribuciones límites que sean utilizados según el caso, alguna de las distribuciones teóricas que se comentan más adelante.

Distribución muestra

El objeto que manejamos en este caso es una distribución de diferencia de medias muestrales, por lo tanto la población de referencia es el conjunto de todas las posibles diferencias de medias $\bar{X}_1 - \bar{X}_2$ de tamaños específicos n_1 y n_2 y no las poblaciones originales de referencia a las que nos referimos cuando hacemos estadística descriptiva.

Distribuciones límites

La distribución normal y la distribución t-student son dos de los modelos teóricos que se ajustan al comportamiento de las distribuciones muestrales de la diferencia de promedios. La normal se utiliza cuando las muestras son suficientemente grandes o cuando se conocen las varianzas de las poblaciones que se quiere comparar. Cuando las muestras son pequeñas y no se conoce la varianza de las poblaciones se utiliza la t-student con un número de grados de libertad que corresponde a la suma del tamaño de ambas muestras menos dos. El concepto de grados de libertad interviene entonces en la determinación de este modelo.

Distribución del estadístico de prueba

La diferencia entre el modelo que se asocia al estimador con respecto al que se asocia al estadístico de prueba subraya una diferencia conceptual que se conecta con el concepto de estandarización. Esto también implica diferencias sutiles, que señalaremos más adelante, cuando se utilizan repre-

sentaciones gráficas de las distribuciones. Formalmente, para tamaños de muestra suficientemente grande y varianzas desconocidas de las poblaciones, el estimador $\bar{X}_1 - \bar{X}_2$ sigue una distribución normal con media $\mu_1 - \mu_2$ y desviación dada por la expresión 3) presentada en el apartado Herramientas. Esto implica que la estadística de prueba asociada a $\bar{X}_1 - \bar{X}_2$ sigue una distribución normal estándar. Casos similares al anterior se aplican al caso de muestras pequeñas cambiando la distribución normal por la distribución t student y teniendo en cuenta el número de grados de libertad.

E. Análisis e interpretación de resultados

Los conceptos que se manejan a nivel de esta unidad de análisis son generales. Por ello es poco lo que se puede decir de manera particular en este tema específico de tests de hipótesis. De todas formas haremos una breve referencia a los conceptos que se ponen en juego en esta unidad.

Nivel de significación estadística

Se define con base en la distribución teórica que se utilice y teniendo en cuenta el establecimiento de una hipótesis nula. Formalmente se puede expresar como $\alpha = P(\text{Rechazar } H_0 \mid H_0 \text{ es cierta})$. En este caso $P(\cdot)$ sería la función de densidad de probabilidad condicionada por la hipótesis nula. Se debe entonces calcular con base en una distribución normal o una t-student.

Tipos de error y potencia

El error de tipo I, se identifica con el nivel de significación que se impone a priori y de manera subjetiva antes de realizar un test estadístico. El error de tipo II, definido como $\beta = P(\text{No rechazar } H_0 \mid H_0 \text{ no es cierta})$, no se puede establecer de manera única a menos que se esté trabajando con hipótesis simples en la hipótesis alternativa. Como cuestión específica a tener en cuenta en caso de que se quiera calcular errores de tipo II o potencias para un contraste dado, está la distribución que sigue la diferencia de medias. En este caso, habría que utilizar una distribución t-student no central.

P-valor

Con respecto a este concepto no hay mucho que decir que tenga relevancia desde el punto de vista particular del tipo de problema que nos ocupa. En todo caso, la definición formal de p-valor para este caso la establecemos como: probabilidad de que $\bar{X}_1 - \bar{X}_2$ tome un valor igual o más extremo (en dirección de la hipótesis alternativa) que el que se obtiene con la muestra de datos cuando la hipótesis nula es verdadera.

ANÁLISIS DEL APRENDIZAJE

Introducción

Esta sección se divide en tres subsecciones. En la primera se describen algunas de las características de los tipos de situaciones problemáticas que consideran inferencias acerca de la diferencia de medias de una población. En la segunda se presentan la situación problemática que se propuso a manera de ensayo piloto, la manera como se aplicó y algunos comentarios relacionados con su evaluación. Finalmente, en la tercera subsección, se presenta la propuesta definitiva, los conceptos y procedimientos que se ponen en juego en dicha propuesta y los errores que pensamos que —a priori— se iban a identificar en los estudiantes. También haremos algunos comentarios acerca del uso de representaciones externas relacionadas con el concepto de nivel de significación.

Situaciones problemáticas: caracterización del tipo de problemas

Como una primera aproximación para analizar los fenómenos que usualmente se ponen en juego en el tema que nos ocupa, se hizo una recopilación de una serie de problemas de diferencia de medias que suelen aparecer en la práctica. Con base en esta recopilación se miraron los constructos hipotéticos utilizados en cada problema, es decir, la idea general del fenómeno a partir del cual se plantea el problema. Los resultados de esta observación sin tener en cuenta la frecuencia con la que se presentó cada constructo, fueron:

- Se utilizan a nivel físico y biológico los siguientes constructos: rendimiento, durabilidad, longitud, crecimiento, consumo, resistencia, contenido de una sustancia, efectividad y precisión. Por otra parte a nivel de constructos sociales y psicológicos se utilizan los siguientes: salarios, nivel académico, producción, inteligencia, memoria, habilidades psicomotoras, capacidades intelectuales, consumo de drogas y calificaciones a nivel de encuestas de opinión.

Por otra parte, la revisión de este tipo de problemas generó tres categorías de tipos de problemas de diferencia de medias de muestras independientes que parecen surgir en la práctica, estas categorías son:

Problemas de diseños estadísticos del tipo grupo control versus grupo experimental. En esta categoría se consideran los problemas donde se plantea formalmente (o se hace alusión a) una investigación en la que se quiere comparar dos grupos que se tratan de igualar en todas las condiciones bajo

las cuales se desarrolla un proceso, excepto en la condición que define la diferencia entre control y experimental. Esta categoría de problemas surge en fenómenos relacionados con investigaciones agropecuarias, en biología, en medicina y en muchas ciencias naturales. También en el área social, el paradigma psicoestadístico tuvo un auge notable. En muchas publicaciones de revistas especializadas se pueden encontrar diseños de este tipo. Sin embargo, ahora este tipo de diseños se presenta con menos frecuencia por la influencia actual de los paradigmas interpretativo y crítico.

Problemas de comparación de dos grupos en dos momentos diferentes de tiempo. En esta categoría se consideran los problemas donde se plantea una comparación de dos estados: uno en el pasado donde la situación reflejaba una diferencia (o no diferencia) entre dos grupos sobre los cuales interesa en el presente, establecer si ha sucedido o no un cambio.

Problemas de comparación de dos grupos en un mismo instante de tiempo.

En esta categoría se consideran los problemas donde se plantea la comparación de dos grupos en un mismo instante de tiempo, sin embargo, a diferencia de la primera categoría en este caso no se establece a priori un diseño, simplemente se comparan los dos grupos en la situación donde surge el constructo de investigación. Desde el punto de vista estadístico son problemas más criticables pues se pueden escapar algunas variables de control que pueden influir de manera indirecta en la diferencia.

Luego de reflexionar acerca del tipo de situación problemática que se les debería plantear a los estudiantes, elegimos un tipo de problema que tocara un tema de ciencias sociales, que hubiera sido comentado en la prensa y que se pudiera clasificar dentro de la tercera categoría: problemas de comparación de dos grupos en un mismo instante de tiempo. Consideramos que este tipo de comparaciones ocurrían con más frecuencia dentro del contexto de las ciencias sociales que las otras dos mencionadas y de manera más natural.

Cuestionario piloto

En un artículo reciente de “El Tiempo” aparece un informe donde se afirma: *las mujeres colombianas siguen siendo discriminadas con respecto a su salario. Aunque por sus niveles educativos y su experiencia, las mujeres deberían recibir un salario promedio mayor que el de los hombres, en realidad las mujeres siguen recibiendo un salario inferior.* En “El Tiempo” se menciona un estudio reciente realizado con base en encuestas, que consideró a las personas labo-

ralmente activas en la ciudad de Bogotá. La siguiente tabla contiene algunos de los resultados publicados acerca de dicho estudio:

	Hombres	Mujeres
Personas encuestados	130	145
Salario promedio	\$500.000	\$470.000
Desviación	\$100.000	\$90.000

Tabla A.

- 1) Con base en lo que fue publicado por “El Tiempo”:
 - a. Defina formalmente el parámetro o los parámetros sobre los cuales interesaría realizar una inferencia estadística si se toma como base la afirmación de “El Tiempo”.
 - b. Determine cuál debe ser realmente la población de estudio si se utilizan los datos de la Tabla A. Explique si es o no necesario revisar la definición dada en la parte A y en caso de que sea necesario, vuelva a formalizar la definición del parámetro o de los parámetros dados.
- 2) Verifique con base en los datos de la Tabla A. que se puede concluir acerca de la afirmación dada en “El Tiempo”. Haga explícito en su proceso:
 - a. Las hipótesis estadísticas (nula y alternativa) que serían más apropiadas para analizar la afirmación de “El Tiempo”.
 - b. Una gráfica donde se represente como deben ser las regiones de rechazo y de no rechazo de la hipótesis nula. También explique en que hechos se basa para plantear su representación gráfica.
 - c. La expresión simbólica de la estadística de prueba que utilice para analizar el problema y los cálculos que es necesario realizar.
 - d. El modelo de distribución probabilística con base en el cual usted puede analizar este problema y por qué debe utilizar dicho modelo y no otro.
- 3) Determine un nivel de significación tal que:
 - a. La hipótesis nula sea rechazada.
 - b. La hipótesis nula no sea rechazada
- 4) Explique qué decisión tomaría con respecto a la afirmación planteada en el problema y justifique por qué.

5) De acuerdo al contexto del problema, interprete lo que significa la decisión tomada.

Metodología de aplicación

Este enunciado fue diseñado para trabajar durante dos sesiones de clase de una hora cada sesión. En la primera sesión se pidió a los estudiantes que resolvieran el problema completo de manera escrita e individualmente y que entregaran la solución al terminar el tiempo programado. Antes de comenzar la siguiente sesión, se revisó la solución dada por cada estudiante y se señaló sobre el papel de cada uno de ellos, los lugares donde ocurrieron errores sin explicar explícitamente cuáles eran los errores. En la siguiente sesión los estudiantes trabajaron en grupos (de dos o tres estudiantes). Para esta sesión, se devolvió a cada estudiante la solución individual de la sesión anterior y se retomó nuevamente el problema del día anterior. Ahora se le pidió a cada grupo que escribiera una solución conjunta. Los estudiantes revisaron las soluciones individuales y discutieron cuáles habían sido los errores cometidos por cada uno de ellos. Se explicó a los estudiantes que este trabajo de grupo tenía dos propósitos: por un lado, procurar que la solución conjunta no contuviera los errores que se cometieron individualmente y por otro, que cada uno de ellos se hiciera más consciente de sus errores.

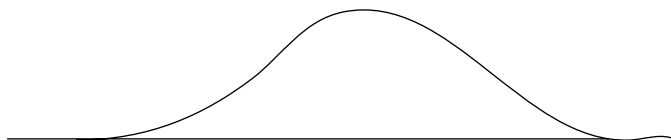
Evaluación

La aplicación de este cuestionario reflejó la gran dificultad que tuvieron muchos estudiantes para responder a la pregunta inicial. El desempeño de los estudiantes en la resolución individual lo resumimos en los siguientes puntos:

- En la primera pregunta (parte A) varios estudiantes (30%) pensaron que el problema sólo involucraba como parámetro la media del salario de las mujeres pero no la media del salario de los hombres. En otras palabras, no consideraban que el problema fuera de comparación de medias.
- En la primera pregunta (parte B) pocos estudiantes (10%) tuvieron presente que la muestra sólo se refería a la ciudad de Bogotá y que la inferencia se pretendía hacer sobre la población colombiana.
- En la segunda pregunta (parte A) se transmite la dificultad planteada en la parte A de la primera pregunta a esta pregunta, pero ahora también se agrega la dificultad del planteamiento de la lateralidad del test (sólo el 40% de los estudiantes plantearon apropiadamente en la hipótesis alternativa el hecho que los hombres recibieran en promedio un salario mayor que el de las mujeres). En este punto nos pareció que los estudiantes confundían lo

que debería ser (que las mujeres recibieran un mejor salario) con lo que se está viendo que es (los hombres reciben un salario mayor que las mujeres). Sin embargo, también había casos donde el test es planteado en forma bilateral (20%), lo cual no apoyaba la hipótesis anterior.

- En la segunda pregunta (parte B) lo más usual fue hacer el dibujo en esta forma



los detalles especificados en cada dibujo variaban de estudiante a estudiante. En general, eran consistentes con los planteamientos bilateral o unilateral (esto se indicaba con una pequeña sombra en las colas), algunos no indicaban números en el gráfico y otros sí, en todos los casos donde fueron indicados los números se entendió que se referían a gráficas estandarizadas.

- En la segunda pregunta (parte C) se presentan en el denominador de la estadística de prueba diferentes versiones de error estándar asociado a la diferencia de medias. Hemos catalogado como correctas las siguientes expresiones para el error estándar de la diferencia de medias:

$$\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \cdot \left(\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\right)} \quad \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Sin embargo, algunos estudiantes presentan la expresión de la derecha usando σ en vez de S y otros presentan la de la izquierda usando $n_1 - 1$ y $n_2 - 1$ en el numerador del factor que corresponde a la varianza ponderada en vez de n_1 y n_2 . Como ya fue dicho en el análisis de contenido, detrás de la escogencia de la estadística de prueba, está presente el tener o no en cuenta el supuesto de homogeneidad de la varianza. Desafortunadamente la formulación del cuestionario no permite que el estudiante exprese si esto lo está teniendo en cuenta o no.

- En la segunda pregunta (parte D) se quería enfatizar en la diferencia que existe entre la estadística de prueba y la distribución de probabilidad que sigue la estadística. La respuesta que dieron la mayoría de estudiantes fue que “el modelo es un z ”. Esto sin embargo, todavía no nos permite distinguir en que piensa real-

mente el estudiante. Puede ser que si esté pensando en la distribución normal en sí misma o puede ser que siga pensando en que el modelo es la estadística de prueba. Por otra parte, realmente la respuesta que normativamente se puede considerar como correcta es que el modelo es una distribución t-student con grados de libertad que varían entre el mínimo de $n_1 - 1$ y $n_2 - 1$ y $n_1 + n_2 - 2$, según la estadística de prueba que haya sido considerada. Este modelo a su vez se puede aproximar a un modelo normal por lo que las muestras son de un tamaño grande.

- En la pregunta 3 (partes A y B), se busca que el estudiante se aproxime al concepto de p-valor. Dentro de los estudiantes que respondieron esta pregunta (25%), sólo uno hace un razonamiento en términos de encontrar el p-valor y decir que para un valor de α mayor que el p-valor se debe rechazar y viceversa.
- La pregunta 4 la responde apenas un 25% de los estudiantes. No se establece una relación entre la decisión y la tercera pregunta, simplemente fijan un nivel de significación (la mayoría lo escoge de 5%) buscan el valor crítico correspondiente y toman la decisión comparando el valor crítico y el que la da la estadística de prueba.
- La pregunta 5 la responden muy pocos estudiantes (10%), posiblemente el tema fue muy largo y no la alcanzan a considerar.
- En el trabajo en grupo los estudiantes logran producir una solución conjunta donde se eliminan la mayoría de los errores señalados por los investigadores. Sin embargo, los errores son corregidos siguiendo un proceso de reflexión superficial. Los estudiantes se limitan a cotejar los temas individuales y a reproducir las partes donde no se han señalado errores para producir la solución conjunta. En otras palabras no se reflexiona sobre los errores cometidos sino se aprovechan los errores no cometidos.

Los resultados anteriores nos obligaron a replantear de nuevo no sólo la formulación de las preguntas sino la metodología de aplicación. A partir de la evaluación anterior y también con base en lo que sucedió en los otros ensayos pilotos y en la primera situación problemática (aplicada en el primer semestre de 1997) se consideró necesario realizar varios ajustes. El cuestionario definitivo quedó como se presenta a continuación.

Propuesta definitiva

Un artículo publicado en “El Tiempo” en octubre de 1996, dice: *En Colombia, por sus niveles educativos y su experiencia, las mujeres deberían recibir un salario promedio mayor que el de los hombres,*

pero en realidad siguen recibiendo un salario inferior. “El Tiempo” también menciona un estudio que consideró personas laboralmente activas de un barrio de Bogotá. La siguiente tabla presenta los resultados publicados:

Salarios	Hombres	Mujeres
menos de 200	3	1
200 a 300	2	3
300 a 400	9	24
400 a 500	55	63
500 a 600	48	46
600 a 700	10	8
700 o más	3	0
Promedio	492,31	470,00
Desviación	102,56	90,22

Tabla B. Distribución de salarios en miles de pesos, de una muestra de personas laboralmente activas, tomada en el barrio Teusaquillo de Bogotá en el mes de octubre de 1996.

¿Realmente las mujeres siguen recibiendo un salario inferior?

1. Tipo de problema, población y parámetro de estudio

- ¿Qué tipo de problema estadístico, reconoce usted en el enunciado anterior?
- Si se utilizan los datos de la tabla de arriba, acerca de que población se puede realmente hacer inferencia.
- Si se toma como base la afirmación de “El Tiempo” (letra itálica), ¿acerca de qué parámetros se debería realizar la inferencia estadística?

2. Análisis de la afirmación

- Suponga que los datos de la tabla de arriba sirven como base para analizar la afirmación mencionada en “El Tiempo”. Plantee las hipótesis nula y alternativa que serían apropiadas para analizar la afirmación.
- Comente cuales son las características de la estadística de prueba que utilizaría para el análisis de la hipótesis, es decir: describa su expresión simbólica, el modelo exacto de la distribución que sigue y la validez de los supuestos para su aplicación.

3. Resultados e interpretación

- a. De acuerdo con los resultados que se obtengan al calcular la estadística de prueba
 - i. ¿Qué nivel de significación es necesario fijar para rechazar la hipótesis nula y cuál para no rechazarla?
 - ii. ¿Qué decisión tomaría y cómo la interpretaría de acuerdo al contexto del problema?
- b. Suponga que el tamaño de las muestras aumentara a 300 hombres y a 300 mujeres, pero que los resultados de los promedios y desviaciones permanecieran iguales. ¿Cómo modifica esto, la decisión tomada en el punto anterior? Explique.

Conceptos y procedimientos que se ponen en juego

En la Tabla 10 que se presenta a continuación se muestran los ítems que a nivel conceptual y procedimental cubre el cuestionario definitivo.

Tabla N° 10.

Conceptos, unidad conceptual y fase involucrada	Procedimientos asociados	Item
Variable Unidad A Fase I	A1. Identificar la(s) variable(s) de estudio.	1a
Muestra Unidad A Fase I	A4. Identificar el número de muestras que se consideran.	1a
	A5. Cuando en el problema se trabaja con dos muestras, identificar si las muestras son independientes o relacionadas.	1a
	A6. Establecer un criterio para determinar si los tamaños de las muestras son “grandes” o “pequeñas”.	2b
	A7. Identificar el tamaño de la muestra.	3a
Población Unidad A Fase I	A8. Establecer cuál es la población de estudio y cuál la de datos.	1b
	A9. Distinguir la población de la muestra.	1b
Nivel de significación Unidad A Fase I	A11. Determinar un nivel de significación para resolver el problema.	3ai
Medidas de resumen estadísticos Unidad A Fase I	A13. Identificar y distinguir los promedios, las proporciones, las desviaciones y/o las correlaciones poblacionales de las muestrales.	3a

Tabla N° 10.

Conceptos, unidad conceptual y fase involucrada	Procedimientos asociados	Item
Parámetros Unidad B Fase I	B1. Identificar o reconocer el o los parámetros de estudio.	1C
Hipótesis de investigación Unidad B Fase I	B3. Identificar la hipótesis de investigación del problema.	2a
	4. Establecer la relación entre la hipótesis de investigación del problema y el parámetro de estudio.	2a
Hipótesis estadísticas Unidad B Fase I	B5. Formular las hipótesis refiriéndose a parámetros de las poblaciones.	2a
Lógica del contraste Unidad B Fase I	B7. Establecer el contraste de hipótesis como un problema de decisión entre dos hipótesis estadísticas.	3aii
	B8. Elegir como hipótesis nula la contraria a la que se desea confirmar.	2a
Varianza del estimador Unidad C Fase II	C3. Identificar el error estándar que se asocia con el estimador.	2b
	C4. Calcular el valor del error estándar.	3a
Estadístico de prueba Unidad C Fase II	C6. Calcular el valor del estadístico de prueba.	3a
Distribución estimador Unidad D Fase II	D1. Identificar una distribución de probabilidad asociada al estimador.	2b
Distribución estadístico de prueba Unidad D Fase II	D4. Identificar una distribución de probabilidad asociada al estadístico de prueba.	2b
	D7. Consultar la tabla de distribución del estadístico para hallar un valor de la distribución dado un nivel de significación.	3a
	D8. Determinar el p-valor de un valor de la estadística de prueba.	3ai
	D9. Establecer criterios de decisión.	3aii
Distribución muestral Unidad D Fase II	D13. Revisar los supuestos que validan la aproximación escogida.	2b

Tabla N° 10.

Conceptos, unidad conceptual y fase involucrada	Procedimientos asociados	Item
p-valor Unidad E Fase III	E1. Interpretar un valor “pequeño” del p-valor como favoreciendo el rechazo de la hipótesis nula.	3ai
	E2. Interpretar un valor “grande” del p-valor como favoreciendo el no rechazo de la hipótesis nula.	3ai
Nivel de significación Unidad E Fase III	E3. Comparar el nivel de significación y el del p-valor para tomar una decisión.	3a
	E4. Comparar el valor crítico asociado al nivel de significación con el valor calculado del estadístico de prueba.	3aii

También se establecieron varias interrelaciones entre los conceptos que se mencionaron en la tabla anterior. Algunas de estas relaciones son citadas en Vallecillos (1996). Ella se refiere a este tipo de relaciones como elementos de significado asociados a un contraste de hipótesis. En la Tabla que aparece a continuación se presentan estas relaciones.

Tabla N° 11.

N°	Interrelaciones	Comentarios
1	En la formulación del contraste de hipótesis se está ante un problema de decisión entre dos hipótesis de las cuales no sabemos cuál es cierta.	Se pone en juego en el punto 3aii y se considera en la conexión entre las unidades B y E.
2	Hay que diferenciar entre hipótesis nula e hipótesis alternativa. Cada una de ellas juega un papel diferente en el contraste.	Se pone en juego en el punto 2a y se considera en la conexión entre las unidades B y C.
3	Hay que diferenciar entre hipótesis unilaterales y bilaterales. El procedimiento de contraste se ve afectado por esta distinción.	Se pone en juego en el punto 2a y se considera en la conexión entre las unidades B y C
4	Existe la posibilidad de cometer un error de tipo I: rechazar una hipótesis nula verdadera.	Aparece de manera implícita en el punto 3aii y se considera en la conexión entre las unidades B y E. La decisión lleva a que si se rechaza se está en posibilidad de cometer error tipo I.

Tabla N° 11.

N°	Interrelaciones	Comentarios
5	El riesgo del error de tipo I viene medido por el nivel de significación.	Aparece de manera implícita en el punto 3a y se considera en la conexión entre las unidades A y E. Se debe escoger el nivel de riesgo de acuerdo al nivel de significación.
6	Existe la posibilidad de cometer un error de tipo II: aceptar una hipótesis nula falsa.	Aparece de manera implícita en el punto 3a _{ii} y se considera en la conexión entre las unidades B y E. La decisión lleva a que si se rechaza se está en posibilidad de cometer error tipo I.
7	Las hipótesis paramétricas se refieren a valores de los parámetros de las poblaciones.	Aparece de manera implícita en los puntos 1c y 2a y se considera en la conexión entre las unidades A y B. Se está pidiendo que se determine a que población realmente se puede inferir y como se plantean las hipótesis relacionadas.
8	No se puede saber con seguridad si se comete uno de los dos tipos de error, ni siquiera una vez tomada la decisión.	Aparece de manera implícita en el punto 3a _{ii} y se considera en la unidad E.
10	El estadístico de prueba es una estandarización del estimador.	Aparece de manera implícita en el punto 2b y se considera en unidad C.
14	La distribución del estadístico depende del valor asociado a parámetros de la población.	Aparece de manera explícita en el punto 2b y se considera en la conexión entre las unidad D y B.
15	El nivel de significación es el área determinada por el valor crítico en la función de densidad de la distribución del estadístico en el muestreo, bajo la hipótesis nula.	Aparece de manera implícita en el punto 3a _i y se considera en la conexión entre las unidad D y B.
16	Un resultado estadísticamente significativo es un resultado cuya probabilidad de ocurrencia, en caso de ser cierta la hipótesis nula es menor que el nivel de significación.	Aparece de manera implícita en el punto 3a _i y se considera en la unidad E.
17	La significación estadística no implica necesariamente la significación (relevancia) desde el punto de vista práctico.	Aparece de manera implícita en el punto 3b y se considera en la unidad E.
18	Un contraste de hipótesis no determina la probabilidad de certeza de H_0 .	Aparece de manera implícita en el punto 3a _{ii} y se considera en la conexión entre las unidades B y E.

Tabla N° 11.

N°	Interrelaciones	Comentarios
20	El parámetro es un valor constante desconocido en la población.	Aparece de manera implícita en el punto 1c y se considera en la conexión entre las unidades A y B.
21	La distribución de un estadístico suele depender del tamaño de la muestra.	Aparece de manera explícita en el punto 2b y se considera en la conexión entre las unidades A y D.
23	El nivel de significación no se determina por la teoría estadística, es fijado por el investigador.	Aparece de manera explícita en el punto 3a y se considera en la conexión entre las unidades A y E.
24	La varianza muestral de un estadístico decrece en función del tamaño de la muestra.	Aparece de manera implícita en el punto 3b y se considera en la conexión entre las unidades A y C.
25	El nivel de significación, junto con las hipótesis nula y alternativa y el estadístico de prueba determinan las regiones crítica y de aceptación en un contraste.	Aparece de manera explícita en el punto 3 y se considera en la conexión entre las unidades A, B C y D.
26	El p-valor del resultado de un estadístico de prueba depende de los datos de la muestra, de la formulación de las hipótesis y de la distribución del estadístico.	Aparece de manera implícita en el punto 3ai y se considera en la conexión entre las unidades A, B, C y E.
29	Un misma hipótesis nula y un mismo nivel de significación pueden determinar diferentes regiones crítica y de aceptación dependiendo del tamaño de la muestra.	Aparece de manera implícita en el punto 3b y se considera en la conexión entre las unidades A, D y E.

Metodología de aplicación

Al igual que en el cuestionario piloto, en la primera sesión se pidió a los estudiantes que resolvieran el problema completo de manera escrita e individual y que entregaran la solución al terminar el tiempo programado. Antes de comenzar la siguiente sesión, se revisaron las soluciones dadas por cada estudiante y se señaló sobre el papel de cada uno de ellos los lugares donde ocurrieron errores. En esta ocasión, igual que en la situación problemática 1, se señalaron los errores con códigos que daban algunos indicios sobre el tipo de error que se había identificado. En la siguiente sesión los estudiantes trabajaron en grupos (de dos o tres estudiantes), se devolvió a cada estudiante la solución individual dada por ellos en la sesión anterior y se retomó el problema propuesto. En esta sesión se pidió a cada grupo que comentarán los errores que habían cometido y que trataran de

explicar por que creían que los habían cometido. Para ello, los estudiantes revisaron las soluciones individuales y discutieron acerca de sus errores.

Errores a priori

A continuación describimos los errores que pensamos que iban a suceder en la aplicación de la propuesta definitiva y que se deducen de la evaluación realizada en la prueba piloto.

Relacionados con la unidad llamada ‘Datos’, pensamos que vamos a encontrar algunos estudiantes que:

- No van a identificar con suficiente precisión la población de estudio en el sentido de la delimitación espacio-temporal.
- No van a identificar que el problema planteado es de comparación de la media de dos muestras independientes.

Relacionados con la unidad llamada ‘Formulación de hipótesis’, pensamos que vamos a encontrar algunos estudiantes que:

- No van a identificar claramente cuál es el parámetro de estudio. Por ejemplo, van a plantear las hipótesis refiriéndose al estimador y no al parámetro.
- No van a reconocer cuál es la lateralidad del test.

Relacionados con la unidad llamada ‘Herramientas’ pensamos que vamos a encontrar algunos estudiantes que:

- No van a calcular correctamente el valor del estadístico de prueba porque no encuentran donde están indicados los tamaños de las muestras o porque no saben cómo calcular el error estándar del estimador.
- No van a identificar correctamente el estadístico de prueba que se utiliza para comparar la media de dos muestras independientes.

Relacionados con la unidad llamada ‘Distribuciones’ pensamos que vamos a encontrar algunos estudiantes que:

- No van a identificar correctamente los grados de libertad. Van a pensar que los grados de libertad están asociados con el número de intervalos de clase de la tablas de frecuencias de los salarios.
- Van a representar de manera contraria a la forma como se planteó la lateralidad en la formulación de hipótesis, las regiones de rechazo y de no rechazo de la hipótesis nula.

Relacionados con la unidad llamada ‘Análisis e interpretación’:

- No van a identificar correctamente el nivel de significación necesario para rechazar o no rechazar la hipótesis nula.
- Van a comparar probabilidades con valores de las distribuciones t o z para tomar una decisión.

Manejo de representaciones externas

A continuación hacemos una revisión acerca de las traducciones entre las diferentes formas de representación externa, en relación con la forma como se pone en juego el concepto de nivel de significación en esta situación problemática. Al igual que en la situación problemática 1, la calculadora gráfica también le aporta al estudiante otro medio de representación de los sistemas de representación de tablas y gráficas de las distribuciones normal y t -student.

1. Traducción de tablas a situaciones (descripciones verbales). En este sentido el punto 3Ai del cuestionario pide explícitamente determinar un nivel de significación para rechazar y no rechazar la hipótesis nula. Para esto el estudiante tienen tres opciones: una, buscar en la tabla del libro, estos niveles y escribir la justificación de lo encontrado en el papel.

La segunda opción, consiste en utilizar el módulo de distribuciones (programa “DIST”) de la calculadora gráfica. En este caso el estudiante debe introducir el valor numérico obtenido en la estadística de prueba y el tipo de distribución apropiado al problema para de esta manera se producir una representación numérica de una probabilidad en la pantalla de la calculadora. La calculadora gráfica le da al estudiante la probabilidad acumulada correspondiente al valor introducido. Entonces, el estudiante debe interpretar si el resultado dado es el que busca o si es necesario realizar alguna operación adicional. En otras palabras el estudiante debe comprender si el resultado numérico que le dio la calculadora corresponde o no con el p -valor.

La tercera opción, consiste en utilizar directamente las tablas de p -valores que vienen dadas en el texto. En este caso, el valor obtenido en la estadística de prueba posiblemente que no se encuentre exactamente en la tabla, entonces el estudiante deberá identificar el calor más cercano al valor obtenido.

En todos los casos, la comprensión del estudiante con relación a la formulación de las hipótesis y su lateralidad en conexión con la representación tabular de la distribución es definitiva para que el estudiante pueda dar una explicación satisfactoria en la representación verbal.

2. Traducción de situaciones (descripciones verbales) a tablas. En esta dirección es como usualmente se plantea la búsqueda del valor crítico correspondiente al nivel de significación dado. En la ‘aproximación clásica’ del

test de hipótesis este tipo de traducción es la que usualmente tiene lugar para determinar las regiones críticas. Sin embargo, intencionadamente quisimos que la situación problemática planteada no propiciara el que el estudiante utilizara la ‘aproximación clásica’. De hecho no se está dando a priori un nivel de significación específico en la situación problemática que se planteó.

3. *Traducción de gráficas a situaciones (descripciones verbales).* En algunas ocasiones la representación gráfica sirve de puente entre la representación tabular y la verbal. Por ejemplo, cuando el estudiante elige la segunda opción de procedimiento mencionada en la traducción tabular a verbal (la primera que se mencionó), recurre a la representación gráfica de la calculadora para tratar de comprender si el valor numérico que ha obtenido corresponde o no con el p-valor.

4. *Traducción de situaciones (descripciones verbales) a gráficas.* El enunciado que se ha dado no propicia que ocurra este tipo de traducción. Sin embargo, algunos estudiantes acompañan su argumentación con la presentación de una gráfica en la que se representa la distribución y el valor del test estadístico. En estos casos, especialmente cuando la representación gráfica contiene errores, el profesor se debe idear maneras de propiciar el que los estudiantes argumenten entre ellos como debería ser una representación gráfica correcta.

5. *Traducción de situaciones (descripciones verbales) a fórmulas.* En el punto 2B del enunciado el estudiante debe dar la representación simbólica del test utilizado.

6. *Traducción de tablas a gráficas.* La calculadora gráfica proporciona un medio ideal para generar representaciones gráficas de los valores de tabulados de una distribución. Por otra parte, su elaboración usualmente hace que el estudiante ponga en juego la comprensión de la noción de escala. Además, la representación gráfica en la calculadora gráfica le aporta al estudiante una manera de verificar si está realizando correctamente la lectura de una tabla de distribuciones.

ANÁLISIS DE ENSEÑANZA

Dentro del currículo del curso el contenido relacionado con esta parte se desarrolló en la quinta semana de clase y luego de ver el tema de inferencia acerca de la proporción de una población. En esta ocasión el profesor se basó únicamente en el texto [1] para la enseñanza.

Comentarios acerca del profesor y el texto utilizado

El texto [1] presenta este tema dentro de un capítulo que titula “Inferencias que involucran dos poblaciones”. Introduce el capítulo con los conceptos de muestras dependientes e independientes y primero aborda el caso de comparación de medias de muestras dependientes antes de entrar al tema objeto de la situación problemática.

Con respecto a la manera como se presentaban los problemas del texto se observaron en cada uno de ellos las siguientes cuestiones:

- Formas como aparece referido el nivel de significación: se pide explícitamente usar un nivel dado, se deja que el estudiante lo elija, o se maneja de otra manera.
- Planteamiento o explicitación de hipótesis que motiven la realización de test de hipótesis: las plantea el problema explícitamente, las deja para que el estudiante las plantee a partir de una hipótesis de investigación, utiliza otra manejo.
- Intención o propósito de la pregunta o preguntas formuladas.
- Presentación de los tests estadísticos.
- Manejo de supuestos que se deben tener en cuenta acerca de las muestras y las distribuciones.

Los resultados de estas observaciones, sin mencionar la proporción de ocurrencia de los tipos de manejo que se presentan, los damos a continuación:

Nivel de significación. hay un porcentaje aproximadamente igual entre los que fijan α de antemano y los que no lo fijan. En muy pocos problemas se presenta una situación donde se pida establecer el nivel de significación de un resultado o donde a partir de un resultado de una muestra, se pida determinar un α de rechazo.

Planteamiento de hipótesis. Se pueden mencionar cuatro posibles maneras de manejar esta parte. Una, la más complicada para el estudiante, es en la que se presenta una hipótesis de investigación; otra, es en la que explícitamente se utiliza una representación simbólica para explicar las hipótesis que se deben contrastar; hay otra, que es similar a la anterior pero en la cual no se utiliza una representación simbólica, solo se utiliza el lenguaje natural; y en la última, se sugiere explícitamente o implícitamente realizar una estimación con base en intervalos de confianza.

De acuerdo con esta notación utilizada en el texto, la formulación estadística que contrasta las hipótesis nula (H_0) y alternativa (H_a), suele considerar típicamente casos tales como:

- $H_0: \mu_1 - \mu_2 = 0$ contra $H_a: \mu_1 - \mu_2 \neq 0$
- $H_0: \mu_1 - \mu_2 = 0$ contra $H_a: \mu_1 - \mu_2 > 0$

- $H_0: \mu_1 - \mu_2 \geq 0$ contra $H_a: \mu_1 - \mu_2 < 0$

Casos donde las formulaciones para la hipótesis nula son del tipo $H_0: \mu_1 - \mu_2 = c$ ($c \neq 0$) se pueden considerar teóricamente pero no es usual tratarlas en la práctica docente. Por otra parte, contrastes del tipo: $H_0: \mu_1 - \mu_2 = 0$ contra $H_a: \mu_1 - \mu_2 = c$ ($c \neq 0$) los cuales se aproximan más a la idea de contraste en el sentido de Neyman y Pearson no ocurren con tanta frecuencia en comparación con los casos señalados inicialmente. Esto refleja que la aproximación al tema se hace bajo la concepción de un test de hipótesis bajo la óptica de Fisher.

Intención de la pregunta. Lo más usual es preguntar si debe o no rechazar la hipótesis nula e interpretar el resultado. Con mucha menos frecuencia se pide: determinar el nivel de significación del resultado o el mínimo nivel de significación de rechazo de la hipótesis nula, o analizar la influencia de los cambios en los tamaños de las muestras en la decisión de un resultado o determinar un tamaño de muestra a partir del cual se empieza a rechazar la hipótesis nula. Por otra parte, hay algunos problemas en donde se sugieren realizar actividades prácticas donde se deben recolectar datos.

Presentación de los tests estadísticos. Aunque a nivel puramente de contenido se podrían considerar a este nivel tests no paramétricos, el programa de contenidos que la Universidad ha propuesto para este curso ha dado tradicionalmente más importancia a las tests paramétricas que se basan en la distribución t-student. Por ello en el diseño no se consideraron algunas de las contrapartes no paramétricas de la t-student para diferencia de medias.

En el texto [1] no se presenta la misma estadística de prueba que se presenta en los textos [2] y [3]. Como ya vimos en el análisis de contenido, hay esencialmente dos tipos de test estadísticos¹ que se podrían asociar a este tipo de problema, uno donde se utiliza como error estándar de la distribución la expresión (1) dada abajo y el otro donde se utiliza la expresión (2).

$$(1) \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \cdot \left(\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\right)} \quad (2) \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Mientras que detrás de la utilización de (1) está implícito el supuesto de igualdad de las varianzas, en (2) no es necesario tener en cuenta este supuesto. Como el profesor había estado trabajando en los cursos anteriores utilizando los otros textos, en los cuales sólo se utilizaba (1), en la primera clase donde se abordó este tema, dijo erróneamente que los grados de libertad del test estadístico dado en el texto era " $n_1 + n_2 - 2$ " cuando en realidad es un nú-

1. Sin considerar los test de tipo no paramétrico.

mero entre el mínimo de “ $n_1 - 1$ ” y “ $n_2 - 1$ ” y el número “ $n_1 + n_2 - 2$ ”. En todo caso el texto no dice como se determina este número y recomienda utilizar como grados de libertad el mínimo entre “ $n_1 - 1$ ” y “ $n_2 - 1$ ”.

Manejo de supuestos. La utilización de (2) en el test estadístico no hace necesario considerar el supuesto de homocedasticidad. Con respecto al supuesto de normalidad que se exige para la convergencia del test estadístico a una distribución t-student, nos llamaron la atención algunos de los problemas propuestos que requieren el uso de MINITAB. Por ejemplo, en los problemas 10.63 o 10.66 (p. 530) se utilizan generadores de números aleatorios para producir distribuciones empíricas de probabilidad de diferencia de medias. Estos problemas sirven para ilustrar la convergencia de la distribución de muestreo de diferencia de medias a la distribución t-student y para mirar lo que sucede con estas distribuciones cuando no se satisface el supuesto de normalidad. Desafortunadamente, en el curso no se contaba con el programa MINITAB y su adaptación a programas como EXCEL o su posible implementación en la calculadora gráfica, demandaba mucho más tiempo del que se tenía disponible.

Características de la aplicación de la situación problemática

Algunas de las características de la instrucción que están presentes en la aplicación de la situación problemática son las siguientes:

- La forma de trabajo que se pide aplicar en la primera sesión es individual y en la segunda sesión es en grupos de tres estudiantes.
- La aplicación en grupo se enfoca directamente en la discusión de los errores que se detectaron en la aplicación individual.
- Los estudiantes han desarrollado mucho más práctica en el manejo de la calculadora gráfica que en el caso de la situación problemática anterior. Por lo menos tenían diez horas de clase en la que ‘jugaban con la calculadora’.
- La calculadora gráfica, al igual que en la primera situación problemática, también se puede utilizar para encontrar y representar de manera gráfica el p-valor o los niveles de significación.
- El profesor no debe intervenir en la aplicación individual. Debe identificar los errores que se presentan en los papeles que le entregan los estudiantes al final de la aplicación individual y producir un reporte de dichos errores para la parte de trabajo en grupo que deberá entregar a los estudiantes.

- El reporte de errores tiene un doble propósito. Por un lado, el de propiciar que los estudiantes reconozcan, en lo posible, las equivocaciones que cometieron en la parte individual y no las repliquen en el trabajo escrito de grupo. Y por otro, hacerle un seguimiento al tipo de errores que comete cada estudiante con fines evaluativos.

EVALUACIÓN

Errores, dificultades y obstáculos

Errores a posteriori

En general, ocurrieron los errores esperados a priori que se habían previsto. Sin embargo, también se identificaron otros errores. A continuación nos referimos a ellos.

Relacionados con la unidad llamada ‘Datos’ se encontró que algunos estudiantes:

- Asocian el tipo de problema que deben resolver con problemas de una sola muestra. Pensábamos que sólo lo iban a confundir con problemas de dos muestras dependientes.

Relacionados con la unidad llamada ‘Formulación de hipótesis’ se encontró que algunos estudiantes:

- Al intentar describir la definición de un parámetro lo que se da son afirmaciones que corresponden a hipótesis.

Relacionados con la unidad llamada ‘Herramientas’ se encontró que algunos estudiantes:

- Confunden la distribución del estadístico de prueba con el estadístico mismo.

Reporte de la lista de errores

Luego de que los estudiantes entregaron las soluciones escritas al problema planteado para trabajar de manera individual, se revisaron las soluciones y se elaboró un reporte para presentarles a los estudiantes. Al día siguiente, se les entregó la descripción de los errores que se habían identificado. A continuación presentamos dicho reporte, tal como se les presentó a los estudiantes:

Errores de los estudiantes en la situación problemática N° 2

- E1: Está mal identificado el tipo de problema.
- E2: Hay imprecisiones en la caracterización del tipo de problema.
- E3: No sabe determinar cuál es la población de estudio.
- E4: No define claramente cuál es el parámetro de estudio.
- E5: Hay un planteamiento incorrecto en las hipótesis estadísticas.
- E6: Está mal identificado el test estadístico o hay imprecisiones en su identificación.
- E7: No identifica correctamente la distribución del estadístico o confunde la distribución del estadístico de prueba con el estadístico mismo.
- E8: Confunden el p-valor con valores de la distribución de la estadística de prueba o estima de manera incorrecta su valor.
- E9: No se determina correctamente el nivel de significación que es necesario para rechazar, o no rechazar, la hipótesis nula.
- E10: Hace una interpretación incorrecta de la decisión que toma con respecto a la hipótesis nula.

Al igual que en la situación problemática anterior, se elaboró una tabla donde está contenida la información por alumno, de los errores que les fueron señalados en su escrito. También están resaltadas en negrilla las tres estudiantes que fueron observadas en detalle en la interacción del trabajo en grupo. A continuación presentamos la tabla mencionada.

Tabla N° 12.

Estudiante	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	Total de errores
Faihan	•			•	•		•		•		5
Hugo		•		•	•	•					4
Andrés			•			•	•				3
Juan	•		•	•	•	•					5
Ximena		•		•			•	•		•	5
Carlos	•		•			•					3

Tabla N° 12.

Estudiante	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	Total de errores
Norma	•				•			•	•		4
Gelber			•		•	•	•		•		5
Angela	•		•				•			•	4
Martha		•					•				2
Natalia		•	•		•		•		•		5
Carolina	•		•	•		•					4
Frecuencia de errores	6	4	7	5	6	6	7	2	4	2	49
Porcentaje	50%	33.3%	58.3%	41.6%	50%	50%	58.3%	16.6%	33.3%	16.6%	

Con referencia a la tabla anterior se encuentra que los errores más frecuente están relacionados con la determinación de la población de estudio (E3) y con la elección de la distribución del estadístico de prueba (E7). También los errores E1, E5 y E6 ocurrieron en un 50% de los estudiantes. Por otra parte, el error E8 ocurrió con una frecuencia menor de lo esperado (16.6%).

Dificultades y obstáculos

Hemos ubicado en el gráfico que representa la estructura conceptual de contenidos del test de hipótesis, los tipos de error que se identificaron en los estudiantes. Esta ubicación de los errores sugiere cuáles son las conexiones entre las unidades conceptuales que están más débiles. En esta situación problemática parece que hay muchas más dificultades para la formulación del problema y para la elección del test estadístico en conexión con los datos.

Pensamos que una de las mayores dificultades, puede estar en la redacción del enunciado. Por ejemplo, para extraer la información que tiene que ver con la población, se alude inicialmente a los hombres y mujeres laboralmente activos en Colombia, sin embargo, en la tabla en la que se presentan los resultados de la recolección de información, se dice que la muestra es de hombres y mujeres laboralmente activos en el barrio Teusaquillo de la ciudad de Bogotá. El estudiante no cae en la cuenta de esto, o si se da cuenta, le produce un conflicto cognitivo la decisión de cuál es la población

sobre la que verdaderamente se puede hacer inferencia con base en la muestra dada.

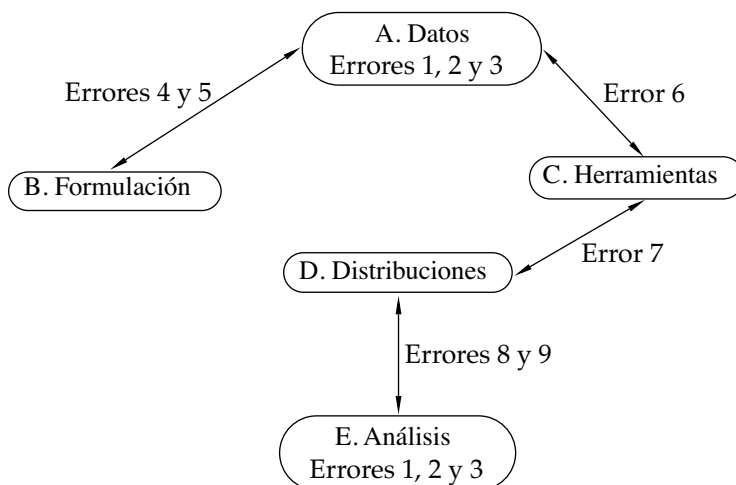


Figura 6.

Una de las razones por la cual se hizo evidente la dificultad para determinar la distribución del estadístico de prueba también está en la forma como se determinaba el tamaño de la muestra. Para varios estudiantes el tamaño de la muestra se deducía a partir del número de categorías de la tabla de frecuencias. Entonces pensaban que el modelo de distribución que se debía utilizar era el t-student con “ $7 + 7 - 2$ ” grados de libertad o con “ $7 - 1$ ” grados de libertad.

Ahora miremos algunos ejemplos concretos de errores de los estudiantes.

Errores 1 y 2. Está mal identificado el tipo de problema o hay imprecisiones en la caracterización del tipo de problema. Carolina escribió que era un problema de análisis de proporciones. Juan escribió: “Test de hipótesis con dos variables”. Martha dijo que era un problema de comparación de dos muestras diferentes y Hugo sólo dijo que era un problema de diferencia de medias. En el caso de Carolina la respuesta coincide con el tema que se ha visto en el capítulo anterior, este tipo de respuesta parece indicar en el estudiante tiene un conocimiento muy pobre del nuevo tipo de problema que se está trabajando. El error de Juan está relacionado con el error 6 y lo vamos a comentar más adelante. La respuesta de Martha y Hugo parece reflejar fallas en diferentes conexiones entre conocimientos. Mientras que la respuesta de Martha no se refiere a que lo que se comparan son las medias y posiblemente asocia la palabra diferente con independencia, la respuesta de Hugo si explicita que se comparan medias, pero no repara en que las mues-

tras deben ser independientes. La respuesta de Carolina y Juan las clasificamos como E1, mientras que las de Martha y Hugo como E2.

Error 3. El estudiante no sabe determinar cuál es la población de estudio. Gelber respondió “se puede hacer inferencia acerca de los salarios”, mientras que Angela dijo “se puede hacer inferencia sobre la población total de los habitantes del barrio Teusaquillo, o sobre la población de Bogotá.” Aunque la respuesta de Gelber no da cuenta de la ubicación espacio-temporal si menciona el tipo de datos de la población: los salarios. La respuesta anterior contrasta con la de Angela. Ella parece ubicar con algo de vacilación la ubicación geográfica pero no explicita el tipo de datos. Aunque ambas respuestas las consideramos como “errores” en realidad reflejan dos formas diferentes e incompletas de conocimiento de lo que significa especificar la población.

Error 4. No se define claramente cuál es el parámetro de estudio. Carolina de manera consistente con la idea de haber expresado que el tipo de problema era de comparación de proporciones, responde así: “P1: las mujeres tienen un nivel de educación más alto y deberían tener un salario promedio más alto. P2: los hombres tienen un nivel de educación más bajo y deberían tener un salario promedio más bajo.” Faihan respondió: “los parámetros son μ_1 y μ_2 .” Juan respondió: “ μ : promedios de los sueldos o salarios recibidos por las mujeres y los hombres.” Es interesante comparar el tipo de conocimiento que se manifiesta en la respuesta de cada uno de estos estudiantes. En el primer caso (el de Carolina), parece que ella piensa que un parámetro debe ser una afirmación donde se exprese una causa y su efecto, también le presta cuidado a señalar la representación simbólica (las letras P1 y P2). Por su parte, Faihan sólo identificó los parámetros con su representación simbólica y no se puede saber si sabría hacer una traducción a la representación verbal. La anterior respuesta contrasta con la de Juan, quien si identifica de qué tratan los parámetros en la representación verbal, pero su representación simbólica refleja, aparentemente, que no tiene clara la traducción de la representación verbal a la representación simbólica o que piensa que el parámetro se resume en una sola letra a pesar de que sean dos grupos.

Error 5. Hay un planteamiento incorrecto en las hipótesis estadísticas.

Para ilustrar este tipo de error es interesante comparar cuatro ejemplos de respuestas. Norma respondió así: “Ho: la cantidad de salario no depende de si usted es hombre o mujer y Ha: la cantidad de salario depende de si usted es hombre o mujer.” Juan respondió así: “Ho: $\mu_{\text{mujeres}} \geq 492.31$ y Ha: $\mu_m < 492.31$ ”. Gelber respondió así: “Ho: $S_H - S_M \geq 0 \Rightarrow S_H > S_M$ (S = sueldo) y Ha: $S_H - S_M < 0 \Rightarrow S_H - S_M < 0$ ”. Y Natalia respondió así: “Ho: $\mu_H \geq \mu_M$ y Ha: $\mu_H < \mu_M < 0$.”

En el primer caso se nota la preferencia por la representación verbal, la ausencia de la palabra promedio y la no determinación de la lateralidad de la prueba. El segundo caso tiene dos aspectos por comentar. En primer lugar, que puede verse como una variante del error en el que representa de manera simbólica el parámetro con la notación que se usa para el estimador (por ejemplo, escribir “Ho: $\bar{x} - \bar{y} \geq 0$ ”). Sin embargo, aunque en este caso la estudiante utiliza la notación “ μ ”, para representar el parámetro ‘media de una población’, no escribe que sea mayor o igual a cero sino que utiliza el valor del promedio muestral del salario de los hombres como referente de comparación. Y en segundo lugar, su representación simbólica sugiere que la hipótesis se centra únicamente en el valor del promedio del salario de las mujeres. El tercer caso, al igual que el primero, no manifiesta que el parámetro sea un promedio; por otra parte, ilustra la dificultad –aparentemente ya superada por el estudiante– para ver como equivalentes dos formas de representación simbólica: la que utiliza la diferencia de los parámetros como mayor o igual a cero y la que compara directamente un parámetro con el otro; finalmente, el estudiante no utiliza la representación simbólica “ μ ”, sino “S” (y explica que “S = sueldo”) lo cual insinúa que él piensa que el parámetro no es la media del salario sino el salario mismo. El último caso ilustra, en concordancia con el uso más extendido en los textos, la forma más frecuente como los estudiantes representan las hipótesis estadísticas; sin embargo, se detecta un error en la forma como se debería plantear la lateralidad del test y no menciona explícitamente en su representación verbal, el significado de la representación simbólica.

Error 6. Está mal identificado el test estadístico o hay imprecisiones en su identificación. Este tipo de error lo podemos ejemplificar con tres casos típicos. Para ilustrar el primero consideramos la respuesta de Juan. En su escrito el dice que el estadístico de prueba es

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Este error es consistente con los errores 4 y 5 identificados anteriormente, pero inconsistente con el error 1. Juan parece que ve este problema como problema de una muestra pero, según el error 1, lo entiende como problema de “dos variables”. Esto sugiere la hipótesis de que Juan tiene una concepción errónea del concepto de variable. El segundo caso lo podemos ilustrar con el error de Gelber que dice que el test estadístico es

$$\frac{\bar{d} - \mu_d}{\frac{S_d}{\sqrt{n}}}$$

Este caso parece similar al anterior. Gelber utiliza el test estadístico que es apropiado para comparar dos muestras relacionadas, sin embargo, a él no se le señaló el error 1. Esto se debió a que dijo lo siguiente acerca del tipo de problema: “es un problema de inferencia que tiene en cuenta dos poblaciones con ~~variables~~ muestras independientes”. Es interesante observar que Gelber, por el tachón de la palabra variables, inicialmente había concebido el problema como de “dos variables”. Por esto uno puede pensar que su concepción inicial acerca del tipo de problema era similar a la de Juan.

Los dos casos comentados anteriormente fueron considerados como errores donde no se identificó el test estadístico. Ahora mostremos un caso donde hay imprecisiones en la identificación del test estadístico. Este es el caso de Andrés que escribió que el test estadístico era

$$\frac{\bar{X}_1 - \bar{X}_2 (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Se detectan dos imprecisiones. Por un lado, la representación simbólica del error estándar hace suponer que se conocen las varianzas de las poblaciones, cuando en realidad son estimaciones basadas en las dos muestras dadas y, por otro lado, en el numerador no aparece el signo de diferencia entre los estimadores y los parámetros. Esta omisión del signo que indica la diferencia entre estimadores y parámetros puede parecer que es un simple descuido, sin embargo, como veremos en el comentario del error 7, la situación de la comprensión de Andrés es mucho más grave.

Error 7. No identifica correctamente la distribución del test estadístico o se confunde la distribución del estadístico de prueba con el estadístico mismo.

En primer lugar comentamos el caso de Andrés. Él escribió el cálculo del test estadístico así:

$$\frac{18.5 + 20,71(492,31 - 470)}{\sqrt{\frac{22,79}{7} + \frac{24,93}{7}}} = -18,29$$

Esta representación numérica del cálculo del test estadístico refleja no sólo que el estudiante piensa que la distribución del estadístico de prueba es una t-student con “7+7-2” o “7-1” grados de libertad sin también una falta de comprensión de muchos otros conceptos o la ausencia total de una traducción de los elementos de la representación simbólica a la representación numérica. En otros casos como el de Ximena, se hace una traducción correcta del la representación simbólica del test estadístico a la representación numérica, sin embargo, ella dijo que la distribución del estadístico era

t-student con 7 grados de libertad y además su representación simbólica utilizaba σ en vez de S, para representar las desviaciones estándar.

Error 8. Confunde el p-valor con valores de la distribución de la estadística de prueba o estima de manera incorrecta su valor. Este error está muy relacionado con el error clasificado como Error 8 en la primera situación problemática. Aunque en esta situación no apareció con la frecuencia esperada, volvió a ocurrir con una mayor frecuencia en el examen final. Este error le fue señalado a Ximena y a Norma. Ximena obtuvo un valor correcto para el test estadístico ($t = 1.906$), sin embargo, hizo el siguiente razonamiento para responder a la pregunta 3Ai: “si $\alpha = 0.05$ y g.l. = 7, el p-valor es 1.89, como $1.906 > 1.89$ se tiene que si a $\alpha = 0.05$ no rechazo H_0 y si $\alpha = 0.025$, el p-valor es 2.36, como $1.906 < 2.36$, se tiene que si a $\alpha = 0.025$ se rechaza H_0 .” El caso de Norma fue diferente. Ella obtuvo un valor incorrecto al calcular el test estadístico, pues le dio $t = 3.871$. Entonces para obtener el p-valor utilizó la aplicación DIST de la calculadora gráfica donde introdujo para d.f. (esto es los grados de libertad con las iniciales en inglés) el valor 273 y para t el valor 3,871, la calculadora da como respuesta la siguiente representación: $P(t) = 0.999932$, entonces Norma dijo que el p-valor era 1. En este caso vemos que Norma no confunde los valores de las distribuciones con las probabilidades de los mismos, pero, no tiene en cuenta que la respuesta que da la calculadora es la probabilidad acumulada hasta el valor 3.871 y no la probabilidad de que t sea mayor que 3.871 que era la probabilidad que ha debido encontrar ($1 - 0.999932$), si no se tuviera en cuenta el error en el cálculo del test estadístico. La forma como procedió Norma se repitió luego con cierta frecuencia en otras situaciones de resolución de problemas en la clase. Vimos que varios estudiantes pensaban que la calculadora les daba directamente el p-valor y no tenían en cuenta el razonamiento que se debía hacer con respecto a la lateralidad del test.

Error 9. No se determina correctamente el nivel de significación que es necesario para rechazar o no rechazar, la hipótesis nula. Dado que la instrucción impartida por el profesor se basó al principio del semestre en el texto [3], los estudiantes no utilizaban con tanta frecuencia la aproximación al p-valor. Por ello no encontramos una relación tan estrecha, como sería de esperar, entre este error y el anterior. Miremos algunos casos. Por ejemplo, Faihan obtiene correctamente el valor del test estadístico ($t = 1.90$) y hace la siguiente representación gráfica

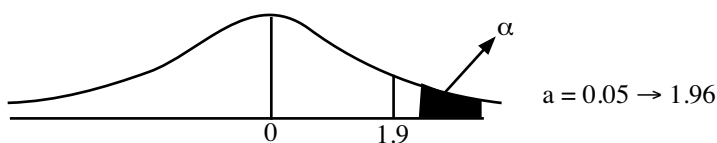


Figura 7.

y luego simplemente escribe: “con $\alpha = 0.05$ se rechaza H_0 [...]”. Este caso nos lleva a pensar nuevamente en el obstáculo que mencionamos en el análisis de la situación problemática anterior, es decir, en la hipótesis de que existe en los estudiantes una noción de comparación de orden restringida a lo puramente lineal, que hace que cuando se ponen en juego en un problema representantes numéricos que corresponden a probabilidades de p-valor o nivel de significación se interpreten como valores sobre la recta real. En el caso anterior es imposible saber si esto sucede realmente, más aun si se tiene en cuenta que Faihan escribe al lado de la gráfica “ $\alpha = 0.05 \rightarrow 1.96$ ”. Sin embargo, en la situación problemática anterior si se hizo explícito argumentaciones del tipo “ $0.05 < 1.9$, entonces rechazo H_0 .”

Error 10. Hace una interpretación incorrecta de la decisión que toma con respecto a la hipótesis nula. Este error le fue señalado a Ximena y a Angela. Sin embargo, se debe tener en cuenta que varios estudiantes no alcanzaron a responder todas las preguntas del cuestionario. En todo caso, como profesores nos sorprendimos por el hecho de encontrar en estas dos estudiantes una interpretación contraria o inconsistente con la decisión tomada. Como investigadores sólo podemos justificar este hecho en términos de la complejidad conceptual que hay detrás de los test de hipótesis. En este caso particular, se evidencia la ausencia de una traducción apropiada entre la representación simbólica que contiene una de las posibles respuestas al problema, “rechazar H_0 ”, y la representación verbal que se debe explicitar en términos del contexto del problema.

10. SITUACIÓN PROBLEMÁTICA N° 3: INFERENCIA ACERCA DEL COEFICIENTE DE CORRELACIÓN DE UNA POBLACIÓN

INTRODUCCIÓN

El tema de esta situación fue el de tests de hipótesis para inferir acerca del coeficiente de correlación (de Pearson) de una población. Se consideraron dos casos: uno, en el que se supone que este coeficiente es igual a cero en la población y el otro, cuando se supone que tiene algún valor diferente de cero en la población.

Al igual que en las dos situaciones anteriores, para la concepción y concreción del diseño se hace un análisis didáctico que contempla tres dimensiones: un análisis del contenido estadístico del tema, un análisis del aprendizaje en el que concretamente se elige un tipo de situación problemática y un análisis de la enseñanza que atiende principalmente al análisis del texto que leyeron los estudiantes antes de enfrentarse al problema propuesto. Al final, se presenta la evaluación de la aplicación de la situación problemática.

Para llegar al diseño de la situación que se puso a prueba se hicieron dos ensayos piloto. El diseño que resultó, difiere con respecto a las situaciones problemáticas N° 1 y N° 2 en el manejo metodológico que se le dio al error. En este caso, se diseñaron dos enunciados en los que explícitamente aparecen señalados varios errores, uno de los enunciados se utilizó para el trabajo individual y el otro para el trabajo en grupo. La tarea del alumno consistió básicamente en explicar y corregir dichos errores.

La forma de aplicación de la situación problemática tuvo tres fases. Primero, los estudiantes explicaron y corrigieron los errores del primer enunciado de manera individual y entregaron sus respuestas por escrito. Luego, los profesores la revisaron las respuestas de los estudiantes y evaluaron si era necesario hacerle algún ajuste al segundo enunciado. Y finalmente, los estudiantes se reunieron de nuevo en la clase, trabajaron en grupos y la tarea que se pidió que realizaran fue esencialmente la misma que la del trabajo individual.

ANÁLISIS DE CONTENIDO

Las principales características de los elementos de contenido se describen a continuación.

A. Datos

Variables

Cuando consideramos una correlación de Pearson, hay presentes dos variables de tipo cuantitativo. Si las variables consideradas están relacionadas se dice que están correlacionadas. En general, el par de variables a las que nos referimos cuando hablamos de una correlación se denotan, en su representación simbólica, como “x” e “y”.

En un análisis de correlación no interesa establecer la relación de causa efecto que pueda existir entre el par de variables. Usualmente un análisis de correlación se hace antes de hacer un análisis de regresión. Es en este último en el que la notación de las variables adquiere el significado en el que se mira “x” como la variable independiente que se puede controlar e “y” como la variable dependiente, cuyo comportamiento se trata de explicar en función de la otra variable. En el análisis de correlación el estatus de “x” e “y” es el mismo, ninguna de ellas se debe suponer como variable dependiente o independiente.

Población

En un análisis de correlación se evidencia la necesidad de distinguir entre lo que es población de estudio y lo que es población de datos. En este sentido, es claro que cuando se quiere hablar de una correlación hay como referencia una población de estudio de la cual que se derivan dos poblaciones de datos, cada una de ellas relacionada con una de las dos variables. Entonces todos los valores posibles de estas dos variables son las que definen las dos poblaciones de datos.

Muestra

La muestra que se toma de las dos poblaciones de datos es pareada, esto significa que de cada unidad de la población de estudio sólo se deriva una pareja de datos. Usualmente este par de datos se obtiene de manera simultánea, esto quiere decir que no hay control experimental sobre ninguna de las dos variables. La muestra simplemente la conforman datos que se suelen denotar en su representación simbólica como parejas de la forma (x_i, y_i) . Por lo tanto el tamaño de la muestra lo determina la cantidad de parejas de datos que se recolectan.

Nivel de significación

Debido a que las distribuciones de probabilidad que usualmente se trabajan en problemas de inferencia acerca de la correlación son continuas se puede establecer a priori niveles de significación dentro de un espectro continuo.

Medidas de resumen estadístico

La r de Pearson es la medida típica de correlación cuantitativa. La correlación muestral se caracteriza porque sólo puede tomar valores entre -1 y 1. El valor -1 indica una correlación perfecta en sentido inverso, mientras que el valor 1 indica una correlación perfecta en sentido directo. Por otra parte, el valor 0 indica la ausencia de correlación. La correlación de Pearson puede decirse que es una manera “estandarizada o relativa” de medir la fuerza y el sentido de una relación. Sin embargo, también tiene su contraparte en términos relativos a la escala de las variables a través de la covarianza muestral.

B. Formulación

Parámetros

El parámetro de estudio es el valor del coeficiente de correlación en una población específica y se denota con la letra griega “ ρ ”. Aunque intervienen dos variables en la definición de este parámetro, debe tenerse en cuenta como se dijo anteriormente, que sólo se considera una población de estudio.

Hipótesis de investigación

En los problemas de correlación la hipótesis que se estudia de manera más natural, es la sospecha de que exista o no una correlación entre el par de variables consideradas. Esto, si se tienen en cuenta la interpretación que se da al valor cero en el coeficiente de correlación, lleva usualmente a plantear un contraste entre una hipótesis donde se supone que no hay correlación y la otra donde se supone que si existe algún nivel de correlación. También es posible considerar problemas donde por ejemplo, se sospeche que existe una alta correlación en sentido positivo.

Hipótesis estadísticas

Las hipótesis estadísticas relacionadas con la correlación puede plantearse, en términos de representaciones simbólicas, en formas tales como: $H_0: \rho = 0$ contra la alternativa $H_a: \rho \neq 0$ (caso bilateral), o $H_0: \rho = k$ (con $k \neq 0$) contra alternativas tales como $H_a: \rho \neq k$ o $H_a: \rho < k$.

Lógica del contraste

En lo que se refiere al planteamiento de la hipótesis estadísticas, vale la pena señalar que si la intención que se tiene es mostrar que existe correlación entre las variables, la lógica del contraste se puede mirar como coherente con la posición o el “sentido de nulidad” que se le da a la hipótesis nula. Sin embargo, si la intención no es mostrar la existencia de correlación, sino más bien de que existe un nivel específico de correlación, ya no

tiene tanto sentido expresar desde el punto de vista conceptual el que la hipótesis nula exprese la no existencia de correlación.

C. Herramientas

Estimador

El estimador de la correlación de la población es r . Su expresión simbólica se representa con expresiones como la siguiente:

$$r = \frac{1}{n-1} \sum \frac{(x - \bar{x})}{s_x} \cdot \frac{(y - \bar{y})}{s_y}$$

El cálculo de r con esta expresión e incluso con expresiones alternativas, implica la realización de una gran cantidad de operaciones aritméticas elementales. Por ello su cálculo exige mayor cuidado y control si se compara con los cálculos que se deberían hacer en las otras dos situaciones problemáticas que fueron consideradas.

La correlación de Pearson es independiente de la posición y de la escala de los datos. Desde el punto de vista conceptual, esto se refleja en su expresión simbólica: debe observarse que se puede definir como el promedio de la suma de los productos estandarizados de las parejas de datos.

Varianza del estimador r

La varianza del estimador bajo el supuesto de que la correlación de la población sea cero es $\sqrt{[(1 - r^2)/(n - 2)]}$. Sin embargo, cuando se consideran situaciones donde la población tenga un coeficiente de correlación diferente de cero, es necesario considerar, para propósitos de inferencia, no el estimador r sino una transformación de dicho estimador. Concretamente se sugiere en la teoría, usar la transformación de Fisher de r dada por $z_r = \text{Log} \left(\sqrt{[(1 - r^2)/(1 + r^2)]} \right)$. En este caso la raíz de la varianza de la distribución de z_r viene dada por $1/\sqrt{(n - 3)}$.

Estadísticos para tests de hipótesis

Con respecto a los tests estadísticos y al igual que en muchos otros tests paramétricos su cociente de estandarización viene dado por la razón entre la diferencia del estimador muestral con las medias poblacionales y su error estándar. De acuerdo con lo discutido anteriormente se pueden presentar dos casos:

- el test estadístico es $t = r / (\sqrt{[(1 - r^2)/(n - 2)]})$ si se supone que $\rho = 0$

- b. Si no se supone que $\rho = 0$ el estadístico es $z = (r_z - \rho_z) / (1 / \sqrt{n-3})$ donde r_z y ρ_z se determinan con base en la transformación de Fisher.

D. Distribuciones

Distribución muestral del estimador r

Los valores de r pueden estar agrupados de diferentes formas. Por ejemplo si la mayoría tiende al valor -1, o tiende al valor de 1 se obtienen distribuciones asimétricas positivas o negativas y hay necesidad de transformar estos datos obteniendo así una distribución no de r sino de r_z que tiene a ser simétrica y normal.

Distribución muestral

Se utiliza la distribución normal cuando hay necesidad de realizar la transformación de Fisher o cuando el tamaño de la muestra es grande, sino se debe utilizar la distribución t-student.

La forma de la distribución es acampanada y tiene un valor esperado igual a cero si la correlación de la población es de cero. En cualquier caso (correlación igual a cero o diferente de cero) los valores del dominio de la distribución estarán entre -1 y 1.

E. Análisis e interpretación de resultados

Los conceptos que se presentan en esta unidad son esencialmente iguales a los que se presentaron en las otras dos situaciones de tests de hipótesis. De todas formas presentamos una breve referencia de estos.

Nivel de significación estadística

Se define con base en la distribución teórica que se utilice y teniendo en cuenta el establecimiento de una hipótesis nula. Formalmente se puede expresar como $\alpha = P(\text{Rechazar } H_0 / H_0 \text{ es cierta})$. En este caso P(l) sería la función de densidad de probabilidad condicionada por la hipótesis nula. Se debe entonces calcular con base en una distribución normal o una t-student.

Tipos de error

Hay dos tipos de errores, el de tipo I o α que consiste en rechazar la hipótesis nula cuando en realidad es verdadera, y el de tipo II o β que consiste en no rechazar la H_0 cuando en realidad es falsa. Es claro que cuanto menor sea el nivel de rechazo, menor será la probabilidad de cometer un error de tipo I y mayor será la probabilidad de cometer un error de tipo II y viceversa.

P-valor

Dado el resultado muestral, el p-valor el mínimo nivel de significación que es necesario establecer para empezar a rechazar la hipótesis nula.

Dependiendo del planteamiento de las hipótesis tiene una definición particular para cuando es de dos colas, o cuando es unilateral a la izquierda o a la derecha.

ANÁLISIS DEL APRENDIZAJE

Introducción

Esta sección se divide en tres subsecciones. En la primera se describen algunas de las características de los tipos de situaciones problemáticas que consideran inferencias acerca de la correlación de una población. En la segunda se presentan las situaciones problemáticas que se propusieron a manera de ensayo piloto, la manera como se aplicaron y algunos comentarios relacionados con su evaluación. Finalmente, en la tercera subsección, se presenta la propuesta definitiva, los conceptos y procedimientos que se pusieron en juego en dicha propuesta y la forma como se implementó su aplicación.

Situaciones problemáticas: caracterización del tipo de problemas

Como una primera aproximación para analizar los fenómenos que usualmente ponen en juego el tema que nos ocupa, se hizo la recopilación de una serie de problemas de inferencia acerca de la correlación de una población que aparecen en la práctica. Con base en dicha recopilación se miraron algunos de los constructos hipotéticos utilizados en cada problema, es decir, la idea general del fenómeno a partir del cual se planteaba el problema.

Como resultado de la revisión hecha, y sin atender a la frecuencia con la que se presentó cada constructo, se concluyó que se utilizan estimaciones de la correlación en campos tales como: psicología, salud, educación, biología, negocios, demografía, lingüística y finanzas, entre otros. Algunos ejemplos particulares asociados con cada uno de los campos mencionados son:

- Relación entre el peso y la estatura (salud).
- Relación entre satisfacción por la realización de un tipo de trabajo y la propensión por dejarlo (psicología).
- Relación entre longitud y edad de algún tipo de planta (biología).

- Relación entre edad de un automóvil y precio del mismo (negocios).
- Relación entre cantidad de población y desempleo (demografía).
- Relación entre la proporción de variaciones sintácticas de cierto tipo y el nivel socioeconómico (lingüística).

Además del criterio basado en las áreas de estudio, también se pueden sugerir otros criterios para clasificar o caracterizar los problemas de correlación. Por ejemplo, se pueden pensar en términos del uso que se da a un coeficiente de correlación. Como sugiere Philips (1992) el coeficiente de correlación se puede utilizar como un criterio de validez o de confiabilidad o como un índice de otro tipo. Otro criterio puede ser en términos del carácter bivariado o multivariado de los estudios relacionados con problemas donde se presentan matrices de correlaciones: Finalmente, también se puede pensar en términos de si el análisis de correlación se mira o no en conjunción con el análisis de regresión. En el último caso se ve la correlación como un índice de causación.

Cuestionario piloto

Para este tema inicialmente realizamos una prueba piloto en el que se planteaba a los estudiantes un contexto que consideraba la utilización de indicadores de tipo demográfico. Además, se quiso que en la situación que se planteara se pudieran comparar más de dos variables con el fin de obtener correlaciones plantear el uso de las matrices de correlación. A continuación presentamos el cuestionario.

Se buscó una situación real en donde se pudiera aplicar el tema de correlación y poder comparar más de dos variables con el fin de obtener correlaciones positivas, negativas y correlaciones nulas.

Primera parte (cuestionario para trabajar de manera individual)

Especialistas en economía y demografía han venido realizando estudios acerca de las tasas de nacimiento, suicidio, crecimiento económico y productividad. Se ha establecido que entre estos tipos de tasas diferentes tipos de relaciones. En un estudio particular algunos especialistas han referenciado las siguientes afirmaciones:

- Afirmación N° 1. En países con un desarrollo tecnológico alto tales como Japón, Estados Unidos, Alemania, Inglaterra, Francia, Italia y Canadá, se tienen bajas tasas de nacimientos (TN) asociadas con altas tasas de suicidio (TS). Algunos demógrafos piensan que, a nivel mundial, la correlación entre estas dos variables es intensa, pero menor que -0.9.

- Afirmación N° 2. Por su parte, algunos economistas afirman que la correlación entre las variables tasa de crecimiento (TC) y tasa de productividad (TP) es mayor a 0.8 ya que en general, independientemente de los países que se estudien, altas tasas de crecimiento están asociadas con altas tasas de productividad.
 - Afirmación N° 3. Tanto economistas como demógrafos, afirman que las tasas de suicidio no parecen estar correlacionadas con las tasas de productividad.
- 1) Plantee las hipótesis estadísticas que se pueden asociar a cada una de las tres hipótesis de investigación mencionadas anteriormente.
 - 2) En la página siguiente se presentan tres gráficas donde se puede verificar a nivel descriptivo si son válidas dichas hipótesis. Desafortunadamente los investigadores no colocaron en el mismo orden las gráficas y olvidaron ponerle algún rótulo de referencia a los ejes. Suponga que en las gráficas se refleja lo que se dice en las afirmaciones. Explique cuál debería ser el nombre de cada gráfica y sugiera un nombre para rotular cada uno de los ejes.

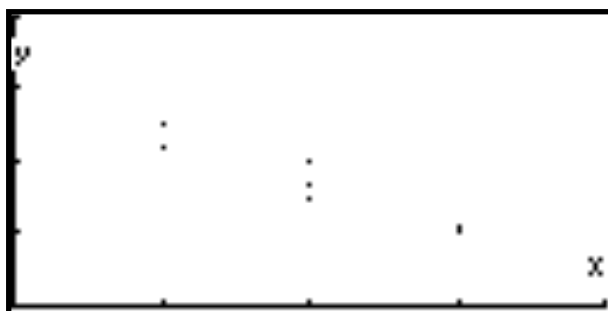
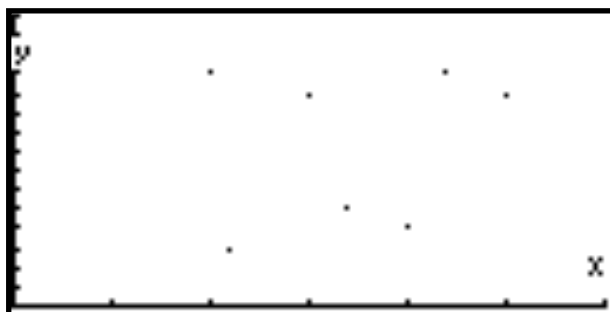
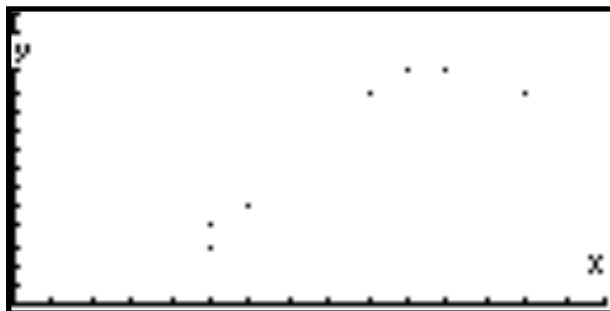
En la tabla siguiente se recopila una muestra de 7 países donde se presentan las tasas de nacimiento, suicidio, productividad y crecimiento.

Tasas	Japón	EE.UU	Alemania	Inglaterra	Francia	Italia	Canadá
TN	3	3	2	2	2	1	1
TS	1	1.1	1.5	1.7	2.0	2.2	2.5
TP	12	3	11	5	4	12	11
TC	10	5	9	6	5	11	13

Tabla A. Tablas de datos

- 3) Compruebe la hipótesis de investigación que planteó en la afirmación N° 1.
- 4) Responda y justifique porqué:
 - a. ¿Con cuál modelo de distribución probabilística trabajaría la hipótesis de investigación que se puede asociar a la afirmación N° 3?
 - b. ¿Cuál es el error estándar que se debería utilizar?
- 5) En otro estudio para analizar la afirmación N° 1 se aumentó el tamaño de la muestra de $n = 7$ a $n = 30$ y se obtuvo el mismo valor para el coeficiente de correlación muestral que se encontró

en el numeral 3) ¿Cómo influye este cambio del tamaño de la muestra en la decisión e interpretación que dio usted como respuesta al problema No 3?



Revisión de las repuestas de los estudiantes y preparación de la segunda sesión

El cuestionario anterior se le dio a los estudiantes para que lo resolvieron en forma individual. Los escritos individuales de los estudiantes se revisaron y se escogieron dentro del conjunto de sus repuestas, algunas donde se reflejaban errores. Entonces, para plantear el trabajo que se debía hacer en grupo durante el segundo día, se diseñó un segundo cuestionario en el que se le pedía a los estudiantes identificar y discutir los errores de cada ítem.

Hubo discusión entre los profesores (investigadores) para ponerse de acuerdo en cuáles de los diferentes errores eran los que se debían poner a discusión de los estudiantes en la siguiente sesión. El enunciado de la segunda parte refleja el acuerdo al que se llegó.

Segunda parte (cuestionario para trabajar en grupo)

- 1) Corrija y explique el error que se presenta en la siguiente respuesta. Este punto se refiere al planteamiento de las hipótesis que corresponden a la afirmación N° 3.

$$H_0: N_1 = N_2$$

N_1 : Tasa de suicidio

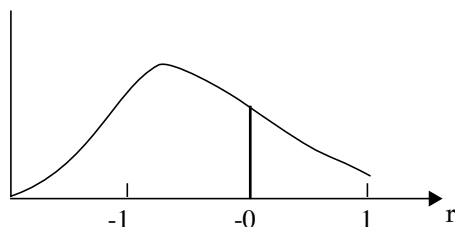
$$H_a: N_1 \neq N_2$$

N_2 : Tasa de productividad

- 2) Con relación al punto 3: corrija y explique los errores que se presentan en la siguiente expresión.

$$Z = \frac{z_r - \rho_r}{\sqrt{\frac{1}{n-3}}} = \frac{1,83 - (-0,95)}{\sqrt{\frac{1}{14-3}}}$$

- 3) Explique cuál es el error en la siguiente gráfica donde se pretende ilustrar la distribución muestral de r si $\rho = -0,9$. Proponga una nueva gráfica donde se ilustre de manera más apropiada la decisión de rechazo o de no rechazo de H_0 .



- 4) Con relación al punto 4, explique cuál es el error en la siguiente argumentación:
 - a. El modelo de la distribución probabilística que se debe utilizar es el "binomial" ya que se trata de afirmar algo sobre la corre-

lación existente entre dos variables y para este tipo de problemas resulta ser el más adecuado.

- b. El E.E. que debe ser utilizado será el mismo que se utiliza en la determinación de los intervalos de confianza:

$$EE_{(X)} = \frac{1}{\sqrt{n-3}}$$

- 5) Con relación al punto 5, explique que está bien y que está mal en la siguiente argumentación:

Si a pesar de incrementar “n” a 30 unidades, el coeficiente de correlación muestral obtenido es igual al obtenido en el problema N° 3, este cambio en el tamaño de la muestra no afecta la decisión, ni la interpretación dadas como respuesta. Lo que si cambiaría sería el

E.E._(X) ya que este es igual a: $\frac{1}{\sqrt{n-3}}$

Evaluación

A nivel investigativo, las explicaciones escritas de los errores no satisficieron las expectativas de los investigadores. Pensábamos que detrás de las explicaciones que explícitamente pedimos a los estudiantes (por escrito), íbamos a identificar de manera más precisa algunas de sus principales dificultades conceptuales. En realidad encontramos en muchos casos, que dichas explicaciones sólo contenían variaciones explicativas de cómo se debía corregir el error y no de porqué se había cometido el error.

A nivel didáctico, se encontró que, en general, los estudiantes lograron presentar correcciones apropiadas a los errores indicados en el cuestionario de la segunda parte. Pensamos que esto se debió principalmente a que en los cuestionarios individuales ya se habían señalado los errores de algunos de los estudiantes. Los resultados de esta experiencia nos ayudaron a apoyar empíricamente nuestra creencia en el hecho de que hacer pensar a los estudiantes sobre sus propios errores les ayuda a que incurran en ellos con menos frecuencia y a que comprendan mejor el tema. Por otra parte, debido a la importancia relativa y diferente que le dábamos cada uno de los profesores a los errores que se habían identificado en los estudiantes, se nos complicó la escogencia de los errores que considerábamos más importantes. Esta experiencia nos obligó a establecer y a desarrollar criterios para seleccionar los errores para poder focalizar la discusión sobre ellos.

Ahora haremos algunos comentarios más puntuales relacionados con el planteamiento y las respuestas de los estudiantes a la aplicación del cuestionario de la segunda parte. En la primera pregunta se pone a discusión de los estudiantes un error relacionado con la formulación de las hipótesis. El error en sí mismo nos motivó pensar en dos posibles explicaciones con respecto a la forma de conocimiento del estudiante en el que se identificó

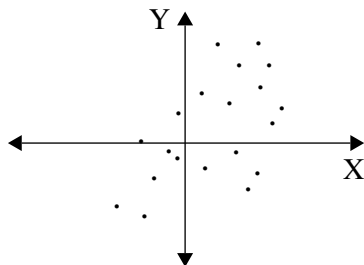
dicho error: desconexión del problema planteado con respecto al significado que se le puede dar al coeficiente de correlación o identificación de los problemas de correlación con problemas de comparación de medias.

Sin embargo, hay un detalle con respecto a la representación simbólica utilizada, que favorece la segunda hipótesis. El estudiante en el que se identificó este error utilizó, para representar los parámetros de las hipótesis estadísticas, una letra más parecida a una “N” mayúscula que a la letra “ μ ” utilizada formalmente para representar la media de una población. Esto corresponde con la forma como el profesor escribía, de manera descuidada, la letra “ μ ” en el tablero (terminaba siendo más parecida a una N mayúscula que a la letra μ).

En la segunda pregunta se ha tomado en el test estadístico el valor n , del tamaño de la muestra, como el total de valores (14) y no como el total de parejas (7). La aparición de este error también se puede relacionar con la hipótesis de identificación de los problemas de correlación con problemas de comparación de medias. Sin embargo, no se pudo establecer una correlación significativa entre los estudiantes en los que se identificó el primer error con respecto a aquellos en donde se identificó este error. La formulación de esta pregunta también contenía otros elementos que generaron discusión entre los estudiantes:

- la transformación de los valores de la correlación al usar la transformación de Fisher.
- el uso de su representación simbólica de las correlaciones transformadas.
- el cuidado con el manejo de la diferencia de las correlaciones en el test estadístico dado que el signo de la correlación muestral era negativo.

En la tercera pregunta se presenta la gráfica de una distribución muestral de correlaciones cuyos valores deberían estar entre -1 y 1. Una de las repuestas dadas por los estudiantes a esta pregunta fue que la gráfica debía ser de la siguiente forma:



Esta respuesta ilustra una de las formas como se le puede sacar provecho al análisis de los errores. Aunque el estudiante respondió mal, la indagación acerca de este error nos permitió identificar otra concepción errada que no habíamos tenido en cuenta y que en este caso se hizo evidente: algunos estudiantes piensan que la distribución muestral de r se puede ilustrar con un gráfico de dispersión de los datos de la muestra, cuando en realidad con este gráfico sólo se puede ver de manera informal el nivel de correlación de los datos de la muestra.

En el enunciado de la formulación de la cuarta pregunta se manifiesta un hecho recurrente que ya se había detectado en la situación problemática anterior y en las dos primeras preguntas del cuestionario de la segunda parte de este ensayo piloto. Se trata de la presencia de respuestas erradas de los estudiantes cuyo contenido se relaciona con temas que se habían tratado con anterioridad. En el caso de las dos primeras preguntas, comentadas unos párrafos más atrás, se puede hacer referencia a los problemas de diferencia de medias. Y en el caso de la cuarta pregunta, se puede hacer referencia a los problemas de proporción.

Con respecto al contenido mismo, las diferentes respuestas de los estudiantes pusieron en evidencia la complejidad de la elección de un modelo de distribución para hacer inferencia acerca de la correlación. Pensamos que la comprensión del estudiante, se vuelva más compleja por los siguientes hechos:

- El cambio que se debe considerar en el modelo de distribución cuando en la hipótesis nula se debe asumir un valor para la correlación diferente de cero.
- La introducción de la transformación de Fisher, que introduce un nuevo ingrediente en el concepto de distribuciones muestral, que ya de por sí es complejo.
- La identificación asintótica del modelo t-student con el modelo normal. Lo que queremos decir con esto, es que el estudiante en general no logra discriminar apropiadamente entre: utilizar el modelo t-student para problemas donde se supone bajo la hipótesis nula que la correlación es cero y utilizar la distribución normal para problemas donde se debe recurrir a la transformación de Fisher cuando se supone bajo la hipótesis nula que la correlación es diferente de cero. La confusión entonces se genera cuando en los enunciados de los problemas se dan muestras grandes, pues debido a que la distribución t-student converge asintóticamente hacia la normal, se confunden los dos tipos de problemas.

A los dos primeros hechos nos hemos referido implícitamente en los comentarios de las preguntas anteriores. Con respecto al tercero, pensamos que

este hecho hace que se confundan no sólo las representaciones simbólicas con las que se hace referencia a las estadísticas de prueba (hecho que también se presentó en la situación problemática anterior), sino también los argumentos que debían exponer los estudiantes para justificar la razón por la cual se debía usar el modelo t-student o el normal.

En la quinta pregunta se plantea un ejemplo del tipo de argumentaciones que dan algunos estudiantes cuando se trata de que analicen los efectos que genera el cambio en el tamaño de la muestra, en la decisión acerca de las hipótesis estadísticas. El error que se plantea en este enunciado llama la atención porque se puede deber a que se esté confundiendo el valor del estimador con el del test estadístico, pues al estudiante le parece natural que se cambie el error estándar y no se modifique nada. Otra cuestión que también ilustra lo que se comentó en el párrafo anterior es la identificación que hace el estudiante con el error estándar $\frac{1}{\sqrt{n-3}}$, cuando la muestra es grande a pesar de que normativamente se deba utilizar el otro error estándar. Finalmente, otro detalle que nos llamó la atención fue la representación “E.E.(x)” que utilizó para representar el error estándar. No pudimos encontrar algún tipo de explicación que justificara el uso de “(x)” en forma de subíndice.

Propuesta definitiva

Para el caso de este diseño se decidió cambiar el estilo del planteamiento de los enunciados. Algunas de las razones que motivaron este cambio fueron: en los cuestionarios individuales varios estudiantes no alcanzaban a responder las últimas preguntas y no se podía recolectar suficiente información con respecto a algunos temas relacionados con la unidad de análisis e interpretación de resultados. Por ello, ahora se plantearon dos enunciados en los que se planteaban soluciones ficticias de un estudiante que contenían errores. La tarea de los estudiantes era identificar en que consistía cada error y corregirlo. En esta ocasión también se dedicaron dos días, uno para trabajo individual y el otro para trabajo en grupo. Los enunciados que se entregaron a los alumnos se presentan a continuación.

Tema A: Avaluadores de finca raíz (aplicación individual)

Una compañía de avaluadores de finca raíz está realizando estudios para decidir cuales son las variables que deberían tener en cuenta si quieren invertir en la compra de apartamentos. Inicialmente avaluaron cinco apartamentos e investigaron en ellos las variables: área en metros cuadrados y edad del apartamento en años. Los avaluadores piensan que estas variables están correlacionadas negativamente, incluso afirman que entre ellas existe una correlación menor que -0.8. Luego de analizar los datos de los cinco

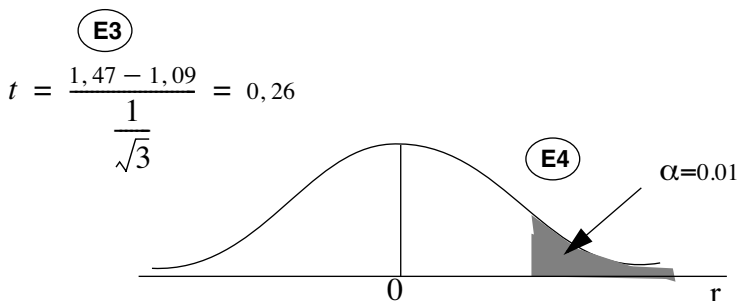
apartamentos obtuvieron un coeficiente de correlación de -0.9. ¿Corrobora este resultado lo que piensan los evaluadores?

Como solución a este problema un estudiante realizó el siguiente test de hipótesis:

Para este problema voy a utilizar un modelo t- student con n-2 grados de libertad ya que las muestras son pequeñas

$$r_z = 1.472$$

$$\rho_z = 1.099$$



Usando un α de 1% encontré un valor en la tabla de 2.575. (E5),

Al comparar el valor calculado con el p-valor de 0.26 que es 0.5941 se tiene que 0.5941 es menor que 2.575. (E6),

Por lo tanto rechazo la hipótesis nula, es decir, existe una correlación mayor que -0.8. (E7),

Las hipótesis estadísticas son $H_0: \rho < -0.8$ y $H_a: \rho > -0.8$

El profesor, luego de revisar esta solución, señaló con las marcas (E1), (E2), etc. los lugares donde el estudiante cometió errores. Usted debe:

- 1. Explicar concretamente en que consistió cada uno de los errores
- 2. Proponer una corrección apropiada

A manera de ejemplo, se ilustra la manera de explicar y corregir dos de los errores cometidos

#	Explicación del error	Corrección apropiada
---	-----------------------	----------------------

E2	Los valores transformados de las correlaciones no tuvieron en cuenta que las correlaciones son negativas	Debemos considerar que si $r = -0.9$ entonces $r_z = -1.472$ y si $\rho = -0.8$ entonces $r_z = -1.099$.
E3	El cálculo del test estadístico que es $Z = \frac{r_z - \rho_z}{\frac{1}{\sqrt{n-3}}}$ no está bien hecho porque no se tuvieron en cuenta los signos de las correlaciones y se dividió por 3 pensando en que el modelo de distribución era el t-student con 3 grados de libertad.	La manera correcta de calcular el test estadístico es $Z = \frac{-1,47 - \langle -1,09 \rangle}{\frac{1}{\sqrt{2}}} = -0,52$

Tema B: Avaluadores de finca raíz (aplicación en grupo)

Una compañía de avaluadores de finca raíz está realizando estudios para decidir cuales son las variables que deberían tener en cuenta si quieren invertir en la compra de apartamentos. Inicialmente avaluaron cinco apartamentos e investigaron en ellos las variables: área en metros cuadrados y edad del apartamento en años. Los avaluadores piensan que estas variables no están correlacionadas. Luego de analizar los datos de los cinco apartamentos obtuvieron un coeficiente de correlación de 0.8. ¿Corrobora este resultado lo que piensan los avaluadores?

Para analizar este resultado un estudiante realizó el test de hipótesis que se presenta a continuación. El profesor, luego de revisar esta solución, señaló con las marcas (E1), (E2), etc. lugares donde el estudiante cometió errores. Usted debe escribir por separado:

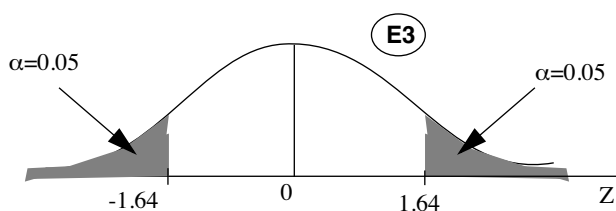
- 1. La explicación tan concreta como se pueda de cuál fue el error que se cometió y
- 2. Proponer una corrección apropiada.

En este problema no es necesario transformar pues queremos averiguar si la correlación es significativa. Aplicaremos directamente el modelo z. (E1),

Las hipótesis son $H_0: r = 0$ y $H_a: r > 0.8$ (E2),

El test estadístico es

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0,8}{\sqrt{\frac{1-(0,8)^2}{3}}} = 2,309$$



El p-valor de 2.309 es 0.947956, este valor es muy grande y **E4** es mayor que un α de 0.05, entonces rechazamos la hipótesis nula y pensamos que no existe correlación entre las variables mencionadas **E5**. Para confirmar lo anterior también vemos que 2.309 es mayor que 1.64 por lo tanto también rechazamos la hipótesis nula **E6**.

Conceptos y procedimientos puestos en juego en la propuesta definitiva

En la Tabla N° 13 que aparece a continuación se presentan los conceptos y procedimientos que se ponen en juego para resolver los enunciados de las propuestas definitivas.

Tabla N° 13.

Conceptos, unidad conceptual y fase involucrada	Procedimientos asociados	Error asociado en los enunciados
Variable Unidad A Fase I	a1. Identificar la(s) variable(s) de estudio.	E1
	a2. Distinguir entre una variable independiente y una dependiente.	E2
	a3. Distinguir entre una variable continua y una discreta.	E1, E2

Tabla N° 13.

Conceptos, unidad conceptual y fase involucrada	Procedimientos asociados	Error asociado en los enunciados
Muestra Unidad A Fase I	a4. Identificar el número de muestras que se consideran. a5. Cuando en el problema se trabaja con dos muestras, identificar si las muestras son independientes o relacionadas. a6. Establecer un criterio para determinar si los tamaños de las muestras son “grandes” o “pequeñas”. a7. Identificar el tamaño de la muestra.	E1 E1, E3 E1, E8 E8
Población Unidad A Fase I	a9. Distinguir la población de la muestra.	E5, E8
Nivel de significación Unidad A Fase I	a10. Identificar el nivel de significación dado en un problema. a11. Determinar un nivel de significación para resolver el problema. a12. Trasladar un valor porcentual a uno proporcional o viceversa.	E4 E4 E2
Medidas de resumen estadísticos Unidad A Fase I	a13. Identificar y distinguir los promedios, las proporciones, las desviaciones y/o las correlaciones poblacionales de las muestrales.	E2
Parámetros Unidad B Fase I	b1. Identificar o reconocer el o los parámetros de estudio. b2. Distinguir un parámetro de su estimador o del valor de una estimación.	E8 E3 y E8
Hipótesis de Investigación Unidad B Fase I	b3. Identificar la hipótesis de investigación del problema. b4. Establecer la relación entre la hipótesis de investigación del problema y el parámetro de estudio.	E8 E8

Tabla N° 13.

Conceptos, unidad conceptual y fase involucrada	Procedimientos asociados	Error asociado en los enunciados
Hipótesis estadísticas Unidad B Fase I	b5. Formular las hipótesis refiriéndose a parámetros de las poblaciones.	E8
	b6. Establecer en las hipótesis estadísticas los valores posibles de los parámetros.	E8
Lógica del contraste Unidad B Fase I	b7. Establecer el contraste de hipótesis como un problema de decisión entre dos hipótesis estadísticas.	E8
	b8. Elegir como hipótesis nula la contraria a la que se desea confirmar.	E8
	b9. Establecer antes de observar los resultados del contraste, las hipótesis.	E8
Estimador Unidad C Fase II	c1. Calcular valores de los estimadores.	E2
	c2. Ubicar valores de las estimaciones en un gráfico.	E4
Varianza del estimador Unidad C Fase II	c3. Identificar el error estándar que se asocia con el estimador.	E3
	c4. Calcular el valor del error estándar.	E3
Estadístico de prueba Unidad C Fase II	c5. Distinguir entre el estimador y el estadístico de prueba.	E3
	c6. Calcular el valor del estadístico de prueba.	E3
Distribución estimador Unidad D Fase II	d1. Identificar una distribución de probabilidad asociada al estimador.	E1, E4
	d2. Identificar los parámetros asociados a la distribución.	E2
	d3. Representar de manera gráfica la distribución del estimador.	E4

Tabla N° 13.

Conceptos, unidad conceptual y fase involucrada	Procedimientos asociados	Error asociado en los enunciados
Distribución estadístico de prueba Unidad D Fase II	d4. Identificar una distribución de probabilidad asociada al estadístico de prueba.	E1, E4
	d5. Identificar los parámetros de la distribución.	E2
	d6. Representar de manera gráfica la distribución del estadístico.	E5
	d7. Consultar la tabla de distribución del estadístico para hallar un valor de la distribución dado un nivel de significación.	E5
	d8. Determinar el p-valor de un valor del test estadístico.	E5
	d9. Establecer criterios de decisión.	E6
	d10. Representar gráficamente las regiones de rechazo y de no rechazo de la hipótesis nula.	E4
	d11. Representar de manera gráfica el p-valor.	E5
Distribución muestral Unidad D Fase II	d12. Distinguir entre la distribución muestral y el modelo de distribución teórico que aproxima dicha distribución.	E1
p-valor Unidad E Fase III	e1. Interpretar un valor “pequeño” del p-valor como favoreciendo el rechazo de la hipótesis nula.	E5
	e2. Interpretar un valor “grande” del p-valor como favoreciendo el no rechazo de la hipótesis nula.	E5, E6
Nivel de significación Unidad E Fase III	e3. Comparar el nivel de significación y el del p-valor para tomar una decisión.	E5, E6
	e4. Comparar el valor crítico asociado al nivel de significación con el valor calculado del estadístico de prueba.	E5, E6
Tipos de error Unidad E Fase III	e5. Identificar el tipo de error que se puede cometer.	E7

También se establecieron varias interrelaciones entre los conceptos que se mencionaron en la tabla anterior. En la Tabla que sigue presentamos estas relaciones. La columna titulada “Tipo de conexión” explicita la conexión que se puede establecer entre cinco unidades conceptuales de contenido que se establecieron en el análisis de contenido.

Tabla N° 14.

N°	Interrelaciones	Tipo de conexión	Explicación
2	Hay que diferenciar entre hipótesis nula e hipótesis alternativa. Cada una de ellas juega un papel diferente en el contraste.	B y C	Aparece explícitamente en los enunciados cuando se afirma que entre las variables existe correlación.
3	Hay que diferenciar entre hipótesis unilaterales y bilaterales. El procedimiento de contraste se ve afectado por esta distinción.	B y C	Aparece explícitamente en uno de los enunciados que dice que la correlación debe ser menor que -0.8, y en el otro en el que dice que las variables no están correlacionadas
6	Existe la posibilidad de cometer un error de tipo II: aceptar una hipótesis nula falsa.	B y E	Aparece implícitamente en planteamiento de los errores E5 y E7
7	Las hipótesis paramétricas se refieren a valores de los parámetros de las poblaciones	A y B	Aparece explícitamente en los enunciados
10	El estadístico de prueba es una estandarización del estimador	C	Aparece explícitamente en los enunciados
14	La distribución del estadístico depende del valor asociado a parámetros de la población.	D y B	En estos casos depende en particular del valor que se supone para la correlación de la población.
16	Un resultado estadísticamente significativo es un resultado cuya probabilidad de ocurrencia, en caso de ser cierta la hipótesis nula es menor que el nivel de significación.	E	Aparece explícitamente en el planteamiento de los errores E6, E4
21	La distribución de un estadístico suele depender del tamaño de la muestra.	A y D	En uno de los enunciados como son 5 datos utiliza la t-student
23	El nivel de significación no se determina por la teoría estadística, es fijado por el investigador.	E y A	No ocurre en los enunciados

Tabla N° 14.

N°	Interrelaciones	Tipo de conexión	Explicación
25	El nivel de significación, junto con las hipótesis nula y alternativa y el estadístico de prueba determinan las regiones crítica y de aceptación en un contraste.	AB y CD	Aparece explícitamente en el planteamiento de los errores E4, E8, E2 y E3.
26	El p-valor del resultado de un estadístico de prueba depende de los datos de la muestra, de la formulación de las hipótesis y de la distribución del estadístico.	AB y CE	Aparece explícitamente en el planteamiento de los errores E4 y E6.
28	Un misma hipótesis nula y un mismo nivel de significación pueden determinar diferentes regiones crítica y de aceptación dependiendo de la lateralidad de la prueba.	BD y E	Aparece explícitamente en el planteamiento del error E4 del segundo enunciado.
29	Un misma hipótesis nula y un mismo nivel de significación pueden determinar diferentes regiones crítica y de aceptación dependiendo del tamaño de la muestra.	AD y E	Aparece explícitamente en el planteamiento del error E3 del segundo enunciado.

Metodología de aplicación

En la primera sesión se pidió a los estudiantes que resolvieran el problema en forma individual y que entregaran la solución por escrito al terminar el tiempo programado. Aunque en esta ocasión, también se revisaron las respuestas individuales no se le entregaron a los estudiantes para la sesión en grupo. En la siguiente sesión los estudiantes trabajaron en grupos (de dos o tres estudiantes) y se les indicó que entregaran una solución escrita por grupo.

Errores a priori

Como ya se había mencionado, los errores detectados en las situaciones problemáticas anteriores nos sirvieron para el diseño de los enunciados. Debido a la estrategia metodológica que se siguió, en esta ocasión no tenía sentido postular errores a priori. En todo caso, el diseño en sí mismo refleja algunos de los errores que a priori hubiéramos esperado. Entre otros aparecen los siguientes: tener dificultad al plantear las hipótesis, reemplazar valores de manera equivocada en el test estadístico cuando el coeficiente de correlación es negativo, tener en cuenta únicamente el tamaño de la muestra para utilizar un test t o z , conclusiones erradas al interpretar el resultado

del test y hacer gráficos donde no se indican apropiadamente sus elementos.

Manejo de representaciones externas

A continuación hacemos una revisión acerca de las traducciones entre las diferentes formas de representación externa, en relación con la forma como se pone en juego el concepto de nivel de significación en esta situación problemática. Al igual que en las situaciones problemáticas anteriores, la calculadora gráfica también le aporta al estudiante otro medio de representación de los sistemas de representación de tablas y gráficos de las distribuciones normal y t-student.

1. Traducción de tablas a situaciones (descripciones verbales). En el error E3 (tema B) y en el error E4 (tema A), la representación gráfica mostrada obliga al estudiante a que consulte las tablas para dar una explicación sobre los errores. Sin embargo, el estudiante también tiene la opción de recurrir a la calculadora gráfica y de esta manera evadir la interpretación de la representación tabular. En los casos de ambos errores, la comprensión del estudiante con relación a la formulación de las hipótesis y su lateralidad en conexión con la representación tabular de la distribución es clave para que el estudiante pueda dar una explicación satisfactoria en la representación verbal.

2. Traducción de situaciones (descripciones verbales) a tablas. Esta traducción está presente en los errores E5 (tema A) y E4 (tema B). En el primer error está presente el uso de la ‘aproximación clásica’ al test de hipótesis, mientras que en el segundo está presente la ‘aproximación al p-valor’. Sin embargo, en el segundo caso pensábamos a priori, que el estudiante iba a preferir el uso de la calculadora gráfica sin recurrir a las tablas.

3. Traducción de gráficos a situaciones (descripciones verbales). La representación gráfica muchas veces sirve de puente entre la representación tabular y la verbal. Sin embargo, en este caso la situación no es así. Ahora las tablas o las representaciones numéricas de la calculadora sirven de puente para dar una explicación acerca de los errores que hay en las representaciones gráficas.

4. Traducción de situaciones (descripciones verbales) a gráficos. En ambos enunciados, este tipo de traducción la debería realizar el estudiante cuando llegara al punto donde debía revisar la consistencia entre lateralidad planteada en la hipótesis de investigación del enunciado y lo que quedaba reflejado en las gráficas. Sin embargo, dado que existe en medio de esta traducción la representación simbólica de las hipótesis, se colocó a propósito en el tema

A la representación simbólica de las hipótesis estadísticas, al terminar la solución del problema.

5. Traducción de situaciones (descripciones verbales) a fórmula. En el tema B aparece explícitamente en el enunciado la representación simbólica del test de hipótesis. En este caso el estudiante debe verificar si lo dicho en el enunciado tiene coherencia con la expresión simbólica del test. Por otra parte, el tema A, la situación que se maneja es diferente pues no aparece de manera explícita el test. En este tema el estudiante tiene una representación numérica de por medio, pero adicionalmente tiene un factor distractor que es el nombre t que se le da a la estadística cuando en realidad es z .

ANÁLISIS DE ENSEÑANZA

Dentro del currículo del curso el contenido relacionado con esta parte se desarrolló en la semana número catorce de clases y luego de ver el tema de inferencia acerca de la independencia de variables categóricas con base en tablas de contingencia. En esta ocasión el profesor se basó en el texto [1] y en el texto [3] para la enseñanza.

Comentarios acerca de los textos utilizados

El texto [1] presenta este tema en el capítulo 13 que titula “Correlación lineal y análisis de regresión”, Sin embargo, presenta un avance de los aspectos descriptivos del coeficiente de correlación en capítulo 3 titulado “Análisis descriptivo y presentación de datos bivariados”. Por contraste, en el texto [3] se presenta seguidos los capítulos donde se trata la parte descriptiva y la parte de inferencia.

Como fortalezas del texto [3] se puede mencionar la construcción conceptual que presenta acerca del coeficiente de correlación (Fernández y Monroy, 1995) y con respecto al texto [1], la presentación de estudios de casos reales. En cuanto a las debilidades del texto [3] se debe subrayar la escasa presencia de problemas que hagan más énfasis en la comprensión conceptual sobre la procedimental, mientras que el texto [3], se puede mencionar como debilidad, la ausencia del tratamiento teórico del caso de inferencia cuando el coeficiente de correlación es diferente de cero.

Con respecto a la manera como se presentaban el tema y los problemas en los textos, se observaron las siguientes cuestiones:

- Proporción de problemas en los cuales explícitamente se debe calcular el coeficiente de correlación a partir de datos dados. Este indicador nos daba una idea aproximada del énfasis que se daba en los problemas a la comprensión conceptual sobre la procedimental.

- La forma como en los problemas se propone el manejo del planteamiento de las hipótesis estadísticas.
- Manejo de las aproximaciones al p-valor y clásica en los ejemplos presentados donde se ilustran soluciones de problemas.

Los resultados de estas observaciones fueron:

Cálculo explícito de r . Hay un porcentaje de 60% en el texto [3] sobre un 48% en el texto [1]. Además, en el texto [1] se presentan casi el doble de ejercicios y problemas con respecto al número que se presenta en el texto [3].

Planteamiento de hipótesis. En el texto [1] se presentaron cinco problemas de nueve en donde el estudiante debe hacer una traducción de lo verbal a lo simbólico. Sin embargo, sólo en dos de estos problemas (ejercicios 13.19 y 13.22) se describe realmente en el texto en forma verbal la hipótesis de investigación, en los otros tres (ejercicios 13.21, 13.23 y 13.27), se pide que se lleve a cabo un test para mirar la significación de la correlación. En los otros cuatro problemas aparece explícitamente una representación simbólica del planteamiento de las hipótesis (ejercicios 13.20, 13.28, 13.29 y 13.30). Por su parte, en el texto [3], en todos los problemas donde se pide realizar un test se debe realizar la traducción de lo verbal a lo simbólico.

Aproximaciones al p-valor y clásica. En el texto [3] no se considera la aproximación al p-valor. En el texto [1], a pesar de que se considera la aproximación al p-valor, en este capítulo, por la forma como se resuelven los ejemplos ilustrativos, el estudiante se ve obligado a utilizar la aproximación clásica. En realidad en el texto [1] se hace una presentación un poco diferente a la del otro texto en el sentido de propiciar la utilización de una tabla que indica de manera directa la significación de los valores de r , en otras palabras el estudiante no tiene necesidad de calcular el test estadístico.

Características de la aplicación de la situación problemática

Algunas de las características de la instrucción que están presentes en la aplicación de la situación problemática son las siguientes:

- Dado que los profesores habían percibido la dificultad de los estudiantes para distinguir entre la explicación del error y su corrección. Para ambas sesiones (individual y de grupo) se entregó una hoja para escribir las respuestas que contenía un formato de dos columnas en el que se indicaba que en la primera

columna se describiera en que consistía el error y en la segunda que se escribiera su corrección.

- A diferencia de lo que ocurrió en las situaciones anteriores, ambas aplicaciones se enfocaron directamente en la discusión y corrección de errores.
- Los estudiantes tienen mucha más destreza en el manejo de la calculadora gráfica que la que tenían en la aplicación de las dos situaciones problemáticas anteriores.
- El profesor no intervino en ninguna de las dos aplicaciones. Hubo dos profesores que actuaron como observadores no participantes. Sin embargo, ellos fueron conscientes del tipo de respuestas que habían dado los estudiantes en la aplicación individual.

Papel de la calculadora gráfica

Cabe anotar que el papel de la calculadora ha sido importante para el desarrollo y claridad del tema ya que en un principio se utilizó para hacer la gráfica del diagrama de dispersión de puntos y para encontrar el valor numérico de r . En esta situación problemática realmente no se necesitaba hacer uso del módulo de estadística de la calculadora gráfica, que entre otras, es muy potente para el análisis de regresión. En realidad debido a la forma en que este se planteó la situación la utilidad se dio en función de la aplicación del módulo de distribuciones de probabilidad. Con éste módulo podían hallar fácilmente p -valores, hacer representaciones gráficas de las distribuciones y verificar si los valores críticos dados en las tabla estaban bien encontrados.

También vale la pena señalar que en todo caso el módulo de estadística se utilizó en las clases para analizar los gráficos de dispersión. En este sentido, su utilización se enfocó en la adquisición de destreza para realizar los “scattergram” y para analizar en las nubes de dispersión generadas las correlaciones entre variables. Por ejemplo, se realizaron actividades en donde se miraba la tendencia a una correlación bien sea positiva o negativa y débil o fuerte.

EVALUACIÓN

El cambio realizado en la metodología de aplicación de esta tercera situación problemática introduce algunas modificaciones en la forma como se va a presentar la evaluación de esta situación, con respecto a la forma como se hizo en las situaciones anteriores.

En primer lugar no se postularon errores a priori. En cambio, se hizo una revisión detallada del historial de errores que se tenía disponible hasta

el momento. Fue realmente con base en este historial que se seleccionaron y diseñaron en los dos enunciados presentados anteriormente los tipos de error que aparecen en éstos.

Por otra parte, hay que señalar que mientras en las dos primeras situaciones el propósito principal se enfocó más en poner en evidencia las concepciones erróneas, en esta situación y quisimos centrar la atención en la indagación sobre las posibles causas de las concepciones erróneas que se habían evidenciado en las situaciones anteriores. Por ello, en este último diseño se enfatizó en que los estudiantes explicaran en que consistían los errores y los corrigieran. Con base en estas explicaciones y correcciones esperábamos obtener más información acerca de estos errores.

Soluciones normativas

Para tener una marco normativo de referencia, los profesores-investigadores preparamos dos soluciones formales que también les entregamos a los estudiantes luego de que respondieran a las dos temas. Las solución normativa desde el punto de vista de los investigadores se presentan a continuación. Para la aplicación individual fue:

Tabla N° 15.

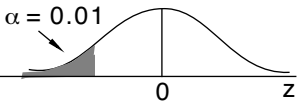
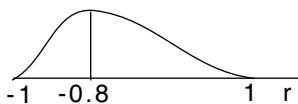
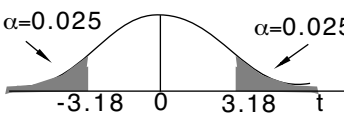
Solución del Tema A	
Explicación del error	Corrección
E1: Modelo de distribución No está correctamente identificado el modelo de distribución del estadístico de prueba.	El modelo correcto de distribución del estadístico de prueba, es el modelo normal.
E4: Lateralidad en la gráfica Según la hipótesis que se está analizando la cola debería ser al otro lado. Por otra parte, no se sabe realmente el nivel de significación que se está utilizando aparece un 0.1 pero luego el estudiante dice que está usando un 1%. Distribución muestral El dibujo realmente no representa la distribución muestral de r (que bajo H_0 , sería sesgada a la izquierda) sino la del modelo normal estándar que aproxima la distribución muestral de los r 's transformados, los r_z 's	Suponiendo que realmente se usa un nivel de significación del 1% la gráfica debería pintarse así:  En caso de que la intención del estudiante fuera representar la distribución muestral de los r 's la gráfica debería ser: 

Tabla N° 15.

E5: Lateralidad en la tabla Buscó el valor en la tabla normal correspondiente a $\alpha=1\%$, como si fuera una prueba bilateral o de dos colas	La prueba es de una cola y la cola es a la izquierda. Por lo tanto el valor que se debe identificar en la tabla es de -2.33.
E6: Aproximación clásica No es necesario calcular el p-valor del resultado del test estadístico, basta comparar el valor obtenido en el test estadístico con el valor de -2.33 correspondiente al nivel de significación.	A. Los valores que se deben comparar son -0.5275 y -2.33
Aproximación al p-valor No era necesario buscar en la tabla normal el valor de -2.33 correspondiente al $\alpha=1\%$. Se debe buscar el p-valor de -0.5275 y compararlo con el de α .	B. El p-valor de -0.5275 es de 0.2989. Este último valor se debe comparar con 0.01
E7: Decisión Si se usa la aproximación al p-valor, no se debe rechazar la hipótesis nula pues el p-valor es mayor que el nivel de significación. Igualmente, si se usa la aproximación clásica, no se debe rechazar la hipótesis nula pues el z calculado para el test estadístico es mayor que el z crítico.	No se rechaza la hipótesis nula pues -0.5275 es mayor que -2.33 (aproximación clásica) No se rechaza la hipótesis nula pues 0.2989 es mayor que 0.01 (aproximación al p-valor).
E8: Formulación de las hipótesis La formulación de las hipótesis se debe plantear al comienzo del problema y no al final. Por otra parte, están mal planteadas las hipótesis, la H_a debería reflejar la hipótesis de investigación y la H_0 negarla.	La formulación de las hipótesis estadísticas debería ser: $H_0: \rho \geq -0.8$ $H_a: \rho < -0.8$

Y para la aplicación en grupo fue:

Solución del Tema B	
Explicación del error	Corrección
E1: Modelo de distribución No está correctamente identificado el modelo de distribución del estadístico de prueba.	El modelo correcto de distribución del estadístico de prueba, es el modelo t-student con $5-2 = 3$ grados de libertad.
E2: Formulación de las hipótesis Están mal planteadas las hipótesis. Se está confundiendo el estimador con el parámetro. Y por otra parte, la prueba es de dos colas y no unilateral derecha como se sugiere al colocar el valor del estimador.	La formulación de las hipótesis estadísticas debería ser: $H_0: r = 0$ $H_a: r \neq 0$

<p>E3: Lateralidad en la gráfica Según la hipótesis que se está analizando la prueba es de dos colas. Por lo tanto el nivel de significación se debe repartir en cada una de los colas. Además el modelo de distribución a que hace referencia la gráfica no es el correcto y por eso el valor crítico tampoco.</p>	<p>Suponiendo que realmente se usa un nivel de significación del 5% la gráfica debería pintarse así:</p> 
<p>E4: Aproximación al p-valor El p-valor no es el número que se obtiene directamente en la calculadora. Debe tener en cuenta que la prueba es de dos colas y la probabilidad complementaria.</p>	<p>El p-valor de 2.3094 es $(1-0.947956)$ multiplicado por dos, es decir 0.104088. Este último valor se debe comparar con 0.05</p>
<p>E5: Decisión e interpretación Al corregir los errores cometidos en la aproximación al p-valor, se debe concluir que no se rechaza la hipótesis nula. Hay contradicción entre la interpretación y el hecho de rechazar si se rechaza H_0 se debería interpretar que existe correlación entre las variables.</p>	<p>No se rechaza la hipótesis nula y se interpreta que no existe correlación entre las variables área en metros cuadrados y edad del apartamento en años.</p>
<p>E6: Aproximación clásica Con la aproximación clásica se deben comparar valores del modelo y no probabilidades. En este sentido no hay error. El error está en basarse en el modelo z para buscar los valores críticos. Esto debe causar un conflicto si no se dan cuenta de este error pues con la aproximación al p-valor no rechazaron y ahora si rechazarían.</p>	<p>Se debe comparar 2.3094 con 3.18, y como 2.3094 es menor que 3,18 se debe concluir que no se rechaza H_0.</p>

Puntuación y análisis

Las respuestas que dieron los alumnos fueron clasificadas dentro de una escala numérica de 0 a 9. Los puntajes se asignaron de acuerdo a los criterios que se describen en la siguiente tabla:

Tabla N° 16.

Criterio	Puntaje
Ni explica ni corrige del error.	0
Identifica y corrige el error en forma muy deficiente.	1
Identifica parcialmente el error pero hace una corrección deficiente.	2
Identifica aceptablemente el error pero su corrección es deficiente.	3

Tabla N° 16.

Identifica muy deficientemente el error y hace una corrección parcial.	4
Explica y corrige parcialmente el error.	5
Presenta una corrección apropiada del error, sin embargo su explicación es deficiente.	6
Explica apropiadamente el error pero lo corrige parcialmente.	7
Explica parcialmente el error y lo corrige apropiadamente.	8
Explica y corrige apropiadamente el error.	9

En la aplicación individual los resultados fueron los siguientes:

	Errores del Tema A (aplicación individual)						
Alumno	E1	E4	E5	E6	E7	E8	Puntaje por alumno
Faihan	7	9	2	7	1	7	33 (61%)
Hugo	8	1	1	5	1	5	21 (39%)
Andrés	7	1	1	1	7	8	25 (46%)
Juan	5	1	1	2	2	2	13 (24%)
Ximena	6	3	9	1	6	9	34 (63%)
Carlos	0	1	1	1	4	6	13 (24%)
Norma	1	5	1	3	1	8	19 (35%)
Gelber	1	4	1	1	1	6	14 (26%)
Angela	4	7	9	2	0	9	31 (57%)
Martha	7	2	1	1	1	2	14 (26%)
Natalia	7	1	1	1	1	2	13 (24%)
Carolina	1	1	2	1	2	6	13 (24%)
	Puntaje global del curso						243 (37.5%)
Puntuación del error	54 (50%)	36 (33%)	30 (28%)	26 (24%)	27 (25%)	70 (65%)	

En la aplicación por grupos los resultados fueron los siguientes:

	Errores del Tema B (aplicación en grupo)						
Alumno	E1	E2	E3	E4	E5	E6	Puntaje por alumno
Grupo 1: Angela, Ximena y Martha	7	2	9	1	2	9	30 (55%)

Alumno	Errores del Tema B (aplicación en grupo)						Puntaje por alumno
	E1	E2	E3	E4	E5	E6	
Grupo 2: Carlos, Andrés y Norma	4	2	4	8	5	1	24 (44%)
Grupo 3: Faihan, Hugo y Juan	5	5	1	1	1	7	20 (37%)
Grupo 4: Gelber, Natalia y Carolina	4	2	7	8	8	4	33 (61%)
Puntaje global del curso							107 (50%)
Puntuación del error	20 (56%)	11 (31%)	21 (58%)	18 (50%)	16 (44%)	21 (58%)	

En la tabla de resultados de la aplicación individual se observa que el desempeño más bajo se dio en lo que se refiere a los errores E6 y E7 y el más alto desempeño se dio con respecto al error E8. A nivel individual sólo dos estudiantes estuvieron por encima de un puntaje superior al 50% y el desempeño global fue del 37.5%. Con respecto a los resultados por grupo se evidencia un mejor resultado global (50%). Por errores, el que más llama la atención por su bajo puntaje es el E2 (31%).

Como veníamos previendo de los resultados de las situaciones anteriores existe en los estudiantes una gran dificultad para manejar los criterios de comparación basados en las aproximaciones clásica y del p-valor, que se refleja en el bajo desempeño individual en los errores E6 y E7.

No esperábamos que el error E8 tuviera un puntaje de desempeño tan alto como el encontrado. Desde nuestro punto de vista, el puntaje alto se debe más a la importancia que se le dio en la puntuación a la identificación por parte de los estudiantes de la mala ubicación de las hipótesis estadísticas (al final de la argumentación) que a la explicación y corrección correcta de la lateralidad expresada en la hipótesis. Por otra parte, los resultados de la prueba en grupos apoyan lo dicho anteriormente. Obsérvese que el puntaje más bajo en dicha prueba, se obtuvo precisamente en E2, que es el error donde se debe discutir el planteamiento de las hipótesis estadísticas.

La comparación de los desempeños individual y de grupos reflejó, como esperábamos, una mejor realización a nivel de grupos. Sin embargo, el mejor desempeño podría atribuirse no a que la interacción en grupos haga posible el obtener mejores puntajes, sino al conocimiento y la experiencia adquiridos en la presentación de la aplicación individual.

Errores dificultades y obstáculos

La revisión de los escritos de los estudiantes reveló la presencia de errores tales como: aplicaciones defectuosas de los algoritmos de aproximación al p-valor o del clásico, decisiones equivocadas de rechazo de la hipótesis

nula e identificación no apropiada de la distribución del test estadístico, entre otros.

Es interesante notar que se presentaron algunos errores que no se habían presentado antes. Entre ellos resaltamos los dos que más nos llamaron la atención:

- Dividir por dos un valor crítico. Por ejemplo si $z = 2.575$ el valor adecuado para la prueba es $2.575/2$.
- El estudiante se guía para el planteamiento de la lateralidad del test, en el resultado muestral y no en lo sugerido por las hipótesis estadísticas.

En todo caso, también encontramos otros errores tales como: confundir el parámetro con variables, confundir el estadístico de prueba con el modelo, no mencionar explícitamente el parámetro que está en juego en el test, tomar el tamaño como el total de los datos como el doble del número de parejas de datos, no transformar los valores de las correlaciones cuando era necesario, plantear las hipótesis con una lateralidad y la gráfica con otra y plantear las hipótesis estadísticas al revés (la nula como alternativa y viceversa) o no explicitar el igual en la hipótesis nula en su forma de representación simbólica.

A continuación presentaremos una selección del tipo de respuestas que dieron los estudiantes que sirven a manera de ilustración de la forma como respondieron los estudiantes.

Ejemplos ilustrativos

Primero vamos a presentar ejemplos del tipo de respuestas dadas por los estudiantes acerca de los errores E1, E4, E5 y E6 del tema A:

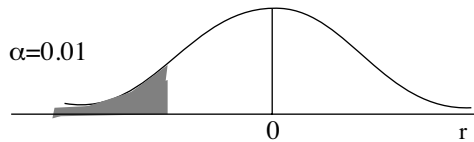
E1. Angela escribió como explicación: “La identificación del modelo t-student no es clara ni la razón por la cual los grados de libertad son $n-2$ ”. Y como corrección, escribió: “Se utiliza un estadístico z , porque los datos se comportan como una normal. No se usan $n-2$ grados de libertad”.

Juan respondió la explicación y la corrección así: “ t no es apropiado ya que si existe un grado de correlación de -0.8 es cercano a -1 por lo cual la distribución es sesgada a la izquierda y tiene que ser normalizada con el método de Fisher, que además utiliza menos 3 para estimar los grados de libertad y no menos dos como es el caso de la t-student”.

En el caso de Angela, se tiene una explicación que consideramos deficiente. Ella identifica que el problema está en la elección del modelo t-student, sin embargo, no discrimina entre los problemas donde se pone a prueba la ausencia de correlación con los que se pone a prueba un nivel de correlación específico. Por contraste la respuesta de Juan, aparentemente, identifica razones que hacen pensar en que él si podría discriminar entre los dos tipos de problemas, pero su corrección es contradictoria pues parece

que sugiere el uso del modelo t-student con $n-3$ grados de libertad y no el modelo normal. Por otra parte, en la respuesta de Angela se puede identificar la aparente falta de discriminación entre lo que es el modelo y lo que es el estadístico.

E4. Andrés explicó y corrigió así: “la gráfica está mal hecha ya que la correlación es negativa, luego el nivel de significación se debería ubicar en la parte negativa así:” (y presenta la gráfica siguiente).

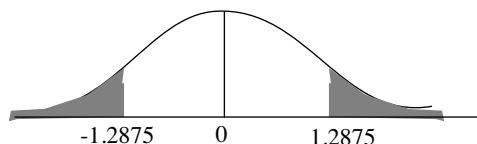


Juan explicó así: “el área sombreada no corresponde a alfa, además es una prueba de dos colas y falta el otro lado” y corrigió así: “el área sombreada corresponde al valor que representa y el $\alpha = 0.005$. (2colas)”.

La explicación de Andrés puede, en principio, parecer que es válida, sin embargo, no se puede determinar si él está pensando en la correlación muestral (-0.9) o en la lateralidad de la hipótesis estadísticas ($H_a: \rho < -0.8$). En realidad, ya vimos como en otras situaciones se identificó que algunos estudiantes tienden a determinar la lateralidad del test con base en la información que aporta la muestra y no la que aportan las hipótesis estadísticas. Con respecto a la respuesta de Juan, parece reflejar la ausencia de conexión entre su respuesta y lo que se insinúa en la hipótesis de investigación.

El otro aspecto que esta presente en este error es el hecho de haber rotulado el eje con la letra r . Con ello se esperaba que el estudiante o la cambiara por z o argumentara alrededor de la distribución muestral. Este detalle fue muy poco identificado por los estudiantes.

E5. Hugo respondió así: “explicación: el valor encontrado en la tabla de 2.575 es correcto pero como el problema y la gráfica no es de una sola cola, el resultado de la tabla es decir el 2.575 se divide en dos lo que da 1.2895 en cada cola”. Y corrigió así: “ $\alpha=1\% \rightarrow \text{TABLA} \rightarrow 2.575 \rightarrow$ como es de dos colas $2.575/2 = 1.2875$. El 1.2875 será el nivel de significación con el cual se rechazaría o no la hipótesis en cada lado de la gráfica” (y presenta la gráfica siguiente).



Este es uno de los errores que más nos sorprendió. La representación gráfica planteada por Hugo sugiere que él ve estos valores como valores del eje. Si suponemos que esto es cierto, entonces concluimos que Hugo piensa en que se puede establecer una relación lineal entre valores de probabilidades (α y p-valor), y los valores de los modelos de distribución (z o t). Así pues, la explicación dada por Hugo da pie para postular la existencia de un obstáculo que consistiría en que se puede establecer probabilidades o valores de los modelos en términos de relaciones lineales.

E6. Carlos explicó y corrigió así: “el p-valor para una distribución normal de 0.5275 es aproximado a 0.199 (además el p-valor no puede ser mayor que 0.5) porque tendremos que tener en cuenta por simetría que unir el valor del otro lado de la tabla.

El valor de 0.5275 proviene de tener en cuenta la corrección que se ilustra en el error E3. Aunque no se pudo establecer de donde sale el p-valor de 0.199 (el valor que se debería encontrar era 0.2989) una posible razón es suponer que el estudiante utilizó el valor para z de 0.26). Sin embargo, es llamativo el hecho de que el estudiante afirme erróneamente que el p-valor no puede ser mayor a 0.5.

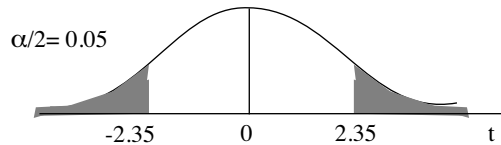
Ahora vamos a presentar algunos ejemplos del tipo de respuestas que dieron los estudiantes del grupo de Angela, Ximena y Martha, acerca de los errores del tema B:

E1. El error consiste en que: “se dice que no es necesario transformar los valores y que utilizará un modelo z por esta razón”. Y corrigieron así: “se debe usar el modelo t-student y por lo tanto no hay necesidad de transformar valores”.

Para la puntuación se había determinado de antemano que el modelo estaría totalmente identificado si además de mencionar que se trataba del modelo t-student, se explicitaban los grados de libertad. Este es el caso de la respuesta anterior, sin embargo, es muy probable que este conocimiento si lo tuvieran los estudiantes. Otro aspecto que llama la atención es el cambio entre antecedente y consecuente en la construcción de la frase “se debe usar el modelo t-student y por lo tanto no hay necesidad de transformar valores” realmente la frase se ha debido construir como: ya que no hay necesidad de transformar los valores entonces debemos utilizar el modelo t-student.

E3. Explicación: “se utiliza $\alpha = 0.1$ en test bilateral y se halla un valor de -1.64. Gráfica mal simbolizada. Se utiliza una gráfica con distribución z”.

Corrección: “el test es bilateral, es decir que $\alpha/2 = 0.05$. valor de la tabla para $t(3, \alpha/2) = 2.35$ ”(y presentan la gráfica siguiente).



A esta explicación y corrección se le asignó el puntaje máximo de 9. Sin embargo, debe observarse que no corresponde con la solución normativa que se había establecido a priori. Esto ilustra las posibilidades, las ventajas y también las desventajas de la utilización de las representaciones gráficas en la instrucción, y en particular, en el diseño de situaciones problemáticas.

E4. Explicación: “el p-valor está mal hallado ya que no se multiplica por 2, al ser un test bilateral” Corrección: “el p-valor se multiplica por 2 en los tests bilaterales $\rightarrow 0.94795 \times 2 = 1.895912$ ”.

E5. Explicación: “el $\alpha = 0.05$ está equivocado en notación. Por la conclusión de rechazar H_0 está equivocado, debido a malos cálculos”. Corrección: “ $\alpha = 0.1$ o $\alpha/2 = 0.05$. No rechazamos H_0 porque el valor de 1.896 es mayor que el α de 0.1. $p\text{-valor} > \alpha \rightarrow$ no rechazo H_0 . En conclusión, se dice entonces que no hay correlación”

La argumentación presentada en los errores E4 y E5, apoya una de las hipótesis mencionadas en los análisis preliminares generales sobre el conocimiento conceptual y procedimental. Por un lado, el procedimental, es claro que se debe multiplicar por dos ya que es un test de dos colas, pero por otro lado, se refleja la debilidad del conocimiento conceptual, por que el estudiante no repara en que el p-valor debe ser un valor que represente a una probabilidad, tranquilamente concluye que el p-valor es 1.895912. Este conocimiento conceptual tal vez le habría podido ayudar al estudiante (en este caso al grupo) a que pudiera caer en la cuenta de que debía hallar el complemento del valor 0.94795 ($1 - 0.94795$) antes de multiplicar por dos.

11. SÍNTESIS Y CONCLUSIONES

A partir de todo el trabajo realizado se generó un esquema general de análisis para llevar a cabo la tarea de diseñar, aplicar y evaluar las situaciones problemáticas. En este esquema de análisis se consideraron varias dimensiones y algunas se trabajaron con más profundidad que otras. Se consideraron entre otros análisis del siguiente tipo:

- Un análisis de algunos aspectos de tipo fenomenológico en el que se intentó identificar el tipo de situaciones reales en las que usualmente se ponen en juego los conceptos relacionados con cada situación problemática.
- Un análisis de contenido en el que se precisan con detalle los conceptos y procedimientos que están presentes en el tipo de tareas en las que se materializaron los diseños de las situaciones problemáticas.
- Un análisis del aprendizaje de los estudiantes que se basó principalmente en el análisis de los errores de los estudiantes, y que dio pie para postular algunos obstáculos de conocimiento de los estudiantes.
- Un análisis de aspectos relacionados con el uso del lenguaje estadístico en el que se consideró el análisis de traducciones entre diferentes sistemas de representación tales como el verbal, el simbólico, el gráfico y el tabular.
- Un análisis de la instrucción en el que se consideró la revisión de los textos que se habían utilizado para la enseñanza durante los últimos semestres y la implementación de la calculadora gráfica.

A continuación vamos a resumir y comentar acerca de las dimensiones que se trabajaron con mayor profundidad.

Conclusiones acerca del análisis de contenido

En primer lugar, se hizo una revisión de los contenidos involucrados en los test de hipótesis con base en la caracterización del conocimiento estadístico en el sentido conceptual y procedimental. En el sentido conceptual, se definieron siguiendo un enfoque sistémico, cinco unidades conceptuales como elementos constitutivos de lo que consideramos un megaconcepto: el test de hipótesis, además se comentaron algunas de las relaciones que se dan entre estas unidades. En el sentido procedimental, se detallaron los diferentes tipos de procedimientos que se ponen en juego alrededor de los conceptos que hacen parte de cada unidad conceptual. Este proceso se hizo de manera general en la parte dos de este documento y luego se replicó de manera particular en cada una de las tres situaciones problemáticas que se

plantearon. En particular, en cada una de las situaciones problemáticas que resultaron como diseños finales, se detallaron los conceptos y procedimientos que en particular se ponían en juego.

Por otra parte, se realizó un análisis en el que se consideraron algunas de las características que se relacionan con el uso del lenguaje estadístico. Concretamente se consideraron algunos de los procesos de traducción entre los diferentes tipos de representación externo en relación con los conceptos de nivel de significación y de p-valor.

También se consideró un modelo de resolución de problemas de tres fases: La primera de comprensión del problema, la segunda de planificación y ejecución y la tercera de evaluación e interpretación. Aunque se identificaron estas fases con el uso de los conceptos y procedimientos involucrados en las unidades conceptuales, esta identificación es muy forzada dado que el uso de algunos de los conceptos y procedimientos asociados a cada unidad conceptual se utilizan necesariamente con la fase con la que fueron identificados.

Como resumen de lo que nos aportó el análisis de contenido podemos mencionar los siguientes puntos:

- Permitió identificar aspectos históricos y epistemológicos que pueden tener incidencia en el aprendizaje y la enseñanza del tema de tests de hipótesis. Por ejemplo, en la teoría de Neyman y Pearson, el enfoque original del test de hipótesis bajo la visión de Fisher, se sustituye por el de la búsqueda de una regla inductiva de comportamiento. Igualmente, con la posición bayesiana se discute la naturaleza subjetiva y objetiva de la probabilidad. En general, se evidencia tanto en el desarrollo histórico, como en la actualidad que los aspectos filosóficos del mismo continúan en debate, en relación con el problema de la inducción (Vallecillos, 1996).
- Establecer un enlace entre la discusión que se reporta en los análisis de aprendizaje con respecto a temas tales como el de errores, dificultades y obstáculos y el de análisis de representaciones externas.
- Un método general para llevar a cabo análisis de contenido de otros temas considerados en los currículos de estadística tal como el de intervalos de confianza.

Conclusiones acerca del análisis del aprendizaje

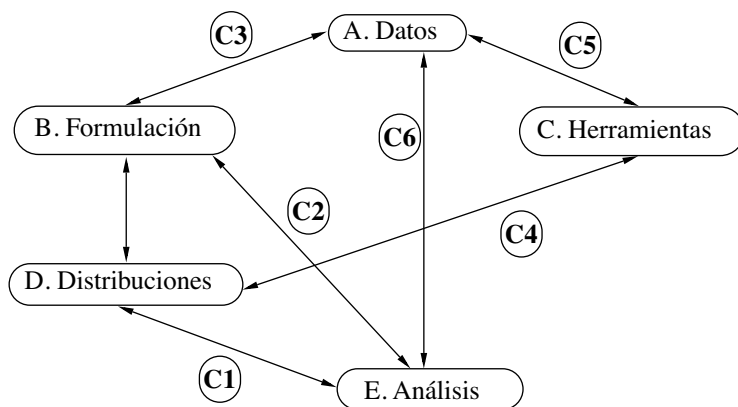
El aprendizaje de los estudiantes fue la segunda dimensión de análisis que se tuvo en cuenta en este estudio. En el marco conceptual se plantearon algunas de las ideas acerca de la comprensión matemática tomadas de Sierpiska (1994), la concepto y concepción tomada de Vergnaud (1993), la de

obstáculos, tomada de Brousseau (1983) que se relacionan con los errores y las dificultades, y la de sistemas de representación de Kaput (1987).

La aplicación de las ideas expuestas en el marco conceptual se materializó en:

- Una descripción general acerca de algunas formas prevalentes del conocimiento estocástico ampliamente referenciadas en la literatura (ver por ejemplo, Shaughnessy (1991) y Scholz (1991).
- Una descripción a priori acerca de los estados hipotéticos de comprensión del tipo de estudiantes considerados en el estudio.
- Una descripción a priori acerca de los errores y las dificultades más comunes del tipo de estudiantes considerados en el estudio.
- Descripciones a posteriori acerca de los errores y las dificultades encontradas en los estudiantes considerados en el estudio, para cada una de las aplicaciones de las situaciones problemáticas.
- Descripciones acerca de las traducciones entre algunas de las formas de representación externa en relación con la aplicación de los conceptos de nivel de significación y p-valor.

Con respecto al análisis de los errores, a continuación presentamos un diagrama en el que se ilustra los lugares donde ubicamos la ocurrencia de la mayoría de errores identificados en la aplicación de las situaciones problemáticas en relación con las conexiones que se pueden establecer con las unidades conceptuales consideradas.



1. En la conexión C1: entre la unidad D y la unidad E

A. Error en la interpretación de formulación de las hipótesis estadísticas.

Este error puede parecer similar al error D que se mencionó en relación con la conexión C3. Sin embargo, al considerarlo como error en esta conexión, se pone el énfasis de la indagación acerca de las dificultades en la interpretación del resultado. Por ejemplo, un resultado significativo lo puede interpretar un estudiante como un resultado que apoya la veracidad de la hipótesis nula.

B. Error en la determinación de la lateralidad de el test. Aunque este error realmente ocurre en la conexión que se establece entre la unidad A y la unidad B (conexión C3), su ocurrencia afecta la interpretación del rechazo o no de una hipótesis nula. Por ejemplo, se puede rechazar una hipótesis nula que no se debía rechazar según el nivel de significación elegido debido a que se propuso una prueba unilateral que debía haberse propuesto como prueba bilateral.

2. En la conexión C2: entre la unidad B y la unidad E

A. Error en la representación del p-valor. Consideramos que ocurre este tipo de error cuando el estudiante hace una inadecuada representación gráfica o simbólica del p-valor. Por ejemplo, en pruebas bilaterales, se representa el p-valor solamente como si el test fuera de una cola.

B. Error en la determinación de valores críticos. Este tipo de error ocurre cuando el estudiante no sabe encontrar el valor crítico correspondiente a un nivel de significación especificado o viceversa. Por ejemplo, en pruebas bilaterales, se determina el valor crítico sin tener en cuenta que se debe dividir por dos el nivel de significación.

C. Error en la determinación del p-valor. Este tipo de error ocurre cuando el estudiante encuentra mal el p-valor correspondiente al valor de un estadístico. Por ejemplo, en una prueba unilateral derecha calcula este valor no como la probabilidad de que el estadístico tome un valor mayor o igual bajo la hipótesis nula, sino al contrario, que tome un valor menor bajo la hipótesis nula.

D. Error en la comparación de valores críticos o del p-valor. Este tipo de error ocurre cuando el estudiante no compara de manera apropiada los valores críticos o los p-valores. Por ejemplo, compara un valor crítico con un p-valor.

3. En la conexión C3: entre la unidad A y la unidad B

A. Error en la identificación del tamaño de la muestra. En algunas ocasiones el estudiante no puede identificar apropiadamente cual es el tamaño de la muestra. Por ejemplo, en casos donde se presenta una distribución de

datos de manera agrupada, el tamaño de la muestra se confunde con el número de categorías de la distribución.

B. Error en la identificación del tipo de problema. El problema que supuestamente está resolviendo un estudiante puede que no coincida con el que normativamente se espera que resuelva. En este caso se dice que el estudiante está identificando mal el tipo de problema que debe resolver. Por ejemplo, el estudiante puede pensar que debe resolver un problema como si las muestras fueran independientes y en realidad las muestras están relacionadas.

C. Error en la identificación del parámetro. Consideramos que ocurre este tipo de error cuando el estudiante no identifica de manera apropiada cual es el verdadero parámetro de estudio. Por ejemplo, se confunde el parámetro con el estimador.

D. Error en la formulación de las hipótesis estadísticas. Consideramos que ocurre este tipo de error cuando el estudiante no identifica de manera correcta si el valor del parámetro debe ser igual, diferente, mayor o menor que cierto valor. Por ejemplo, cuando se tienen hipótesis unilaterales se determina de manera contraria la lateralidad de el test.

4. En la conexión C4: entre la unidad C y la unidad D

A. Identificación incorrecta de la distribución del test estadístico. Ocurre cuando se utiliza un modelo de distribución que no corresponde con el normativamente que se debería utilizar. Por ejemplo, utilizar la distribución normal en lugar de la distribución t-student, cuando en un problema de inferencia acerca de la media de una población se ha estimado la desviación de la misma y la muestra es pequeña.

5. En la conexión C5: entre la unidad A y la unidad C

A. Errores en el cálculo de estadísticos. Son errores que suelen ocurrir cuando se realizan cálculos de estadísticas. Por ejemplo, el cálculo de un error estándar puede resultar errado porque se olvida sacar la raíz cuadrada a una varianza.

6. En la conexión C6: entre la unidad A y la unidad E

A. Errores en la consideración de los supuestos bajo los cuales se realiza el análisis. Se interpretan los resultados como definitivos a pesar de que sólo se satisfacen de manera parcial los supuestos que soportan el procedimiento de contraste. Por ejemplo, cuando no es claro que el muestra haya sido obtenida al azar.

B. Error en la interpretación de la decisión que se tomó. En algunas ocasiones todo el procedimiento de prueba de hipótesis ha sido realizado correctamente pero se interpreta de manera incorrecta el resultado. Por

ejemplo, cuando se interpreta el no rechazo de una hipótesis nula como si en realidad se hubiera rechazado.

Conclusiones acerca del análisis de la enseñanza

Se consideró el análisis de cuatro de los textos de estadística utilizados durante los últimos semestres. Dos de ellos (*Estadística y Sociedad. Conceptos y herramientas básicas del método estadístico.* y *Matemáticas, Azar, Sociedad. Introducción a los conceptos básicos de estadística.*), que ya no eran textos oficiales del programa del curso, influyeron en la enseñanza de los temas que se consideraron en la primera situación y en la tercera. Por otra parte, el texto oficial del curso tuvo su mayor influencia en la aplicación de la segunda situación.

Con respecto a la implementación de las calculadoras gráficas, parece que con ellas se puede favorecer el uso de la aproximación al p-valor sobre la clásica. Sin embargo, es considerable la influencia del tipo de instrucción que se dé en el salón de clase y en particular del tipo de textos que se utilicen. Aunque pensamos que el uso de la representación gráfica (bien sea como medio de verificación de un resultado o como medio de obtenerlo) que proveen las calculadoras gráficas, sólo pudo ser examinado sobre una muestra muy reducida de tres estudiantes, la observación sugiere que los estudiantes usan muy poco el medio de representación gráfica que le da la calculadora. El factor por el cual la calculadora gráfica aparentemente favorece la aproximación al p-valor se debe más a la posibilidad que ofrece la calculadora de tener el resultado numérico de la probabilidad acumulada que a las posibilidades de representación gráfica de la misma.

Con respecto a la evaluación misma de las situaciones problemáticas su análisis puso en mayor evidencia un fenómeno de compartimentalización del conocimiento de los conocimientos conceptual y procedimental y de respuestas que obedecen más al contrato didáctico que al enunciado del problema. También, al interrogar a algunos estudiantes luego de revisar sus respuestas a las situaciones problemáticas se descubren inconsistencias entre la notación empleada y la interpretación dada. Había estudiantes que empleaban correctamente la notación, pero la interpretaban incorrectamente y viceversa, por ejemplo, en la notación de los parámetros.

REFERENCIAS

- Anderson, C.W. y Loynes, R.M. (1987). *The Teaching of Practical Statistics*. New York: John Wiley & Sons Ltd.
- Artigue, M. (1990). Epistémologie et Didactique. *Recherches en didactique des mathématiques*, 10 (2-3), 241-286.
- Artigue, M. (1995). Ingeniería didáctica. En M. Artigue, R. Douady, L. Moreno, y P. Gómez (Eds.), *Ingeniería didáctica en educación matemática* (pp. 33-59). Bogotá: “una empresa docente”.
- Batanero, C., Godino, J.D., Vallecillos, A., Green, D.R. y Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *Int. J. Math. Educ. Sci. Technol.*, 25 (4), 527-547.
- Burriel, G. (1990). Quantitative Literacy: leadership training for masters teachers. En A. Hawkins (Ed.), *Training Teachers To Teach Statistics* (pp. 219-227). London: International Statistical Institute.
- Bisquerra, R. (1989). *Métodos de investigación educativa*. Barcelona: CEAC.
- Butts, T. (1980). Posing Problems Properly. En S. Krulik, (Ed.), *Problem Solving in School Mathematics, 1980 Yearbook*. Reston, VA: NCTM.
- Brewer, J. (1986). Behavioural statistics textbooks: source of myths and misconceptions? En R. Davidson y J. Swift (Ed.), *Actas de la ICOTS II* (pp. 127-131). University of Victoria.
- Brousseau, G. (1983). Les obstacles épistémologiques et les problèmes en mathématiques. *Recherches en didactique des mathématiques*, 4 (2), 164-198.
- Brousseau, G. (1993). Fundamentos y métodos de la didáctica de las matemáticas. En E. Sánchez y G. Zubieta (Eds.), *Lecturas en didáctica de las matemáticas: Escuela Francesa* (pp. 1-67). México: Cinvestav-IPN.
- Centeno, J. (1988). *Números decimales ¿por qué? ¿para qué?* Madrid: Síntesis.
- Chatfield, C. (1988). *Problem Solving. A Statistician's Guide*. London: Chapman and Hall.
- Chevallard, Y. (1985). *La transposition didactique*. Paris: La Pensée Sauvage.
- Chevarney, N., Collier, R. y Fienberg, S. (1977). A framework for the development of measurement instruments for evaluating the introductory statistics course. *The American Statistician*, 31 (1), 17-23.
- Cobb, P. (1993). Characteristics of Classroom Mathematics Traditions: An Interactional Analysis. Conventional Instruction. *American Educational Research Journal*, 29 (3), 573-604.
- Cockcroft, Y. (1985). *Las matemáticas si cuentan. Informe Cockcroft*. Madrid: Ministerio de Educación y Ciencia.
- Douady, R. (1986). Jeux de cadres et dialectique outil-objet. *Recherches en didactique des mathématiques*, 7 (2), 5-31.

- Douady, R. (1995). La ingeniería didáctica y la evolución de su relación con el conocimiento. En M. Artigue, R. Douady, L. Moreno, y P. Gómez (Eds.), *Ingeniería didáctica en educación matemática* (pp. 61-96). Bogotá: una empresa docente.
- Falk, R. (1986). Misconceptions of Statistical Significance. *Journal of Structural Learning*, 9, 83-96.
- Falk, R. y Greenbaum, C. (1995). Significance tests die hard. *Theory and Psychology*, 5 (1), 75-98.
- Fernández, F., Mesa, V., Gómez, P. y Perry, P. (1993). *Estadística y Sociedad*. Bogotá: una empresa docente.
- Fernández, F. (1996). Diseño, desarrollo y evaluación de situaciones problemáticas de estadística (documento de trabajo). Bogotá: “una empresa docente” (<http://ued.uniandes.edu.co/servidor/ued/proyectos/propuestas/proyestad/proyestad.html>)
- Fernández, F. y Monroy, O. (1995). Experiencias en el manejo del coeficiente de correlación de Pearson en un curso de estadística. En P. Gómez, C. Carulla, M. Castro, F. Fernández, C. Gómez, V. Mesa, P. Perry y P. Valero. *Aportes de “una empresa docente a la IX CIAEM* (pp. 107-117). Bogotá: “una empresa docente”.
- Garfield, J. (1995). How Students Learn Statistics. *International Statistical Review*, 63 (1), 25-34.
- Gal, I. y Garfield, J. (1997). Curricular Goals and Assessment Challenges in Statistics Education. En I. Gal y J. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 1-13). Amsterdam: IOS Press Ohmsha.
- Gómez, P., Mesa, V., Perry, P., Fernández, F., Gómez, C. y Marulanda, I. (1993). *Matemáticas y Sociedad*. Bogotá: “una empresa docente” (Documento de trabajo).
- Goetz, J. y Lecompte, M. (1988). *Etnografía y diseño cualitativo en la educación*. Madrid: Mc Graw Hill.
- Grass, R. (1992). L'analyse des données: une méthodologie de traitement de questions de didactique. *Recherches en Didactique des Mathématiques*, 12 (1), 59-72.
- Hacking, I. (1965). *Logic of statistical Inference*. Cambridge: Cambridge University Press.
- Hacking, I. (1995a). *El surgimiento de la probabilidad*. Barcelona: Editorial Gedisa S. A.
- Hacking, I. (1995b). *La domesticación del Azar*. Barcelona: Editorial Gedisa S. A.
- Hawkins, A. (Ed.). (1990). *Training Teachers To Teach Statistics*. London: International Statistical Institute.

- Harwell, M. (1994). Evaluating statistics texts used in education and psychology. Comunicación en la *American Educational Research Association*. New Orleans.
- Hiebert, J. y Carpenter, T. (1991). Learning and teaching with understanding. En D. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 65-100). New York: MacMillan.
- Hiebert, J. y Lefevre, P. (1987). Conceptual and Procedural Knowledge in Mathematics: An Introductory Analysis. En J. Hiebert (Ed.), *Conceptual and Procedural Knowledge: the case of Mathematics*. London: LEA, Publishers.
- Hiebert, J. y Wearne, D. (1987). Procedures Over Concepts: the Acquisition of Decimal Number Knowledge: the case of Mathematics. London. LEA, Publishers.
- Hogg, R. (1990). Statisticians Gather to Discuss Statistical Education. *Amstat News*, 169, 19-20.
- Hogg, R. (1992) Towards Lean and Lively Courses in Statistics. En Florence. S. Gordon and Sheldon. P. Gordon, *Statistics for the Twenty-First Century, Mathematical Association of America*, Series: MMA Notes, 26, 3-13.
- Jaramillo, B. (1989). ¿Será que se puede? En P. Perry, F. Fernández, V. Mesa, y P. Gómez. *Matemáticas, Azar, Sociedad. Una introducción empírica a los conceptos de probabilidad* (pp. 334-350). Bogotá: “una empresa docente”.
- Janvier, C. (1987). Translation Process in Mathematics Education. En C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 27-32). London: LEA, Publishers.
- Kaput, J. (1991). Technology and Mathematics education. En D. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 515-556). New York: McMillan.
- Kaput, J. (1987). Representations Systems and Mathematics. En C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 19-26). London: LEA, Publishers.
- Macnaughton (1996). The Introductory Statistics Course: A new Approach. (<http://www.hookup.net/~donmac>).
- Malara, N. (1989). Probabilità e statistica nella scuola media. Analisi di alcuni libri di testo. Progetto strategico del C.N.R. Università di Modena.
- Mason, J. (1996). Mi comprensión de “la comprensión en matemáticas”. *Revista EMA*, 1, 227-243.
- Moore, D. (1990). Uncertainly. En L.Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 14-25). Washington: The Mathematical Association of America.
- Moore, D. (1992). Teaching statistics as respectable subject. En F. Gordon y S. Gordon (Eds.), *Statistics for the twenty-first Century* (pp. 14-25). Washington: The Mathematical Association of America.

- Morrison, D. y Henkel, R. (Eds.) (1970). *The Significance Test Controversy*. –A reader. Chicago: Aldine.
- NCTM, (1981). *Teaching Statistics and Probability*. Reston, VA: NCTM.
- Oakes, M. (1986). *Statistical inference: a commentary for the social and behavioural sciences*. Chichester: J.Wiley.
- Ortiz, J., Batanero, C. y Serrano, L. (1996). Las frecuencias relativas y sus propiedades en los textos españoles de bachillerato. *Revista EMA*, 2, 19-36.
- Perry, P., Fernández, F., Mesa, V. y Gómez, P. (1989). *Matemáticas, Azar, Sociedad. Una introducción empírica a los conceptos de probabilidad*. Bogotá: “una empresa docente”.
- Perry, P., Fernández, F., Mesa, V. y Gómez, P. (1990). *Matemáticas, Azar, Sociedad. Una introducción empírica a los conceptos de probabilidad*. Bogotá: “una empresa docente”.
- Perry, P., Fernández, F., Mesa, V. y Gómez, P. (1993). *Matemáticas, Azar, Sociedad*. Bogotá: “una empresa docente”.
- Phillips, J. (1992). *How to Think about Statistics*. New York: Freeman and Company.
- Pollard, P. y Richardson, J. (1987). On the probability of making type I errors. *Psychological Bulletin*, 10, 159-163.
- Popper, K. (1967). *La lógica de la investigación científica*. Madrid: Tecnos.
- Radatz, H. (1980). Student’s Errors in the Mathematical Learning Process: a Survey. *For the Learning of Mathematics*, 1, (1), 16-20.
- Rico, L. (1995a). Consideraciones sobre el currículo escolar de matemáticas. *Revista EMA*, 1, 4-24.
- Rico, L. (1995b). Errores en el aprendizaje de la matemáticas. En J. Kilpatrick, P. Gómez y L. Rico. *Educación Matemática* (pp. 69-108). Bogotá: “una empresa docente”.
- Rico, L. (1997). Dimensiones y componentes de la noción de currículo. En L. Rico (Ed.), *Bases teóricas del currículo de matemáticas en educación secundaria*. (pp. 377-414). Madrid: Editorial Síntesis.
- Romberg, T. y Carpenter, T. (1988). Research on teaching and learning Mathematics: Two disciplines of scientific inquiry. En M. Wittrock (Ed.), *Handbook of Research on Teaching* (pp. 850-869). New York: MacMillan.
- Rubin, A. y Rosebery, A. (1990). Teachers’ Misunderstanding in Statistical Reasoning; Evidence from a Field Test of Innovative Materials. En A. Hawkins (Ed.), *Training Teachers To Teach Statistics* (pp. 72-101). London: International Statistical Institute.
- Scholz, R. (1991). Psychological Research in Probabilistic Understanding. En R. Kapadia y M. Borovcnik (Ed.), *Chance Encounters: Probability in Education*. (pp. 213-249). Amsterdam: Reidel.

- Schuyten, G. (1991). Statistical Thinking in Psychology and Education. *Actas de los ICOTS III*. University of Otago, Dudenin (Australia).
- Shaughnessy, J.M. (1991). Research in probability and statistics: reflections and directions. En D. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 465-494). New York: McMillan.
- Sierpinska, A. (1994). *Understanding in Mathematics*. London: The Falmer Press.
- Steinbring, H. (1990). The Nature of Stochastic Knowledge and the traditional Mathematics Curriculum –some experience with in-service training and developing materials. En A. Hawkins (Ed.), *Training Teachers To Teach Statistics*. London: International Statistical Institute.
- Steinbring, H. (1990). The Nature of Stochastic Knowledge and the traditional Mathematics Curriculum –some experience with in-service training and developing materials. En A. Hawkins (Ed.), *Training Teachers To Teach Statistics*. London: International Statistical Institute.
- Tanur, J., Mosteller, F., Kruskal, W., Lehmann, E., Link, R., Pieters, R. y Rising, G. (Eds.). (1989). *Statistics: A Guide to the Unknown*. California: Wadsworth & Brooks.
- Tukey, J. (1977). *Exploratory Data Analysis*. New York: Addison Wesley.
- Vergnaud (1993). La teoría de los campos conceptuales. En E. Sánchez y G. Zubieta (Eds.), *Lecturas en didáctica de las matemáticas: Escuela Francesa*, (pp. 88-117). México: Cinvestav-IPN.
- Vallecillos, A. (1995). Comprensión de la lógica del contraste de hipótesis en estudiantes universitarios. *Recherches en Didactique des mathématiques*, 15, (3), 53-81.
- Vallecillos, A. (1996). *Inferencia estadística y enseñanza: un análisis didáctico del contraste de hipótesis estadísticas*. Granada: Comares.